OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: June 10, 2004, 15:30:58 ; Search time 202 Seconds
                                   (without alignments)
                                   5558.757 Million cell updates/sec

Title: US-09-674-593-10
Perfect score: 246
Sequence: 1 gtcattaacctttgcaagga........ttcatatcaacccacacta 246

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Published Applications NA:*
        1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
        2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
        3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
        4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
        5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
        6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
        7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
        8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
        9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
       10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
       11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
       12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
       13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
       14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
       15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
       16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
       17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
       18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
       19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	C	Score	% Query Match	Length	DB	ID	Description
1	C	38.4	15.6	861	15	US-10-198-846-1685	Sequence 1685, Ap
2	C	36.2	14.7	3588	15	US-10-282-122A-16553	Sequence 16553, A
3		35.6	14.5	507	9	US-09-879-536-826	Sequence 826, App
4		35.6	14.5	24387	10	US-09-764-891-9844	Sequence 9844, Ap
5	C	35.4	14.4	2803	9	US-09-822-849A-479	Sequence 479, App
6	C	35.4	14.4	4480	17	US-10-312-354-67	Sequence 67, Appl
7		34.8	14.1	2264	15	US-10-252-157-264	Sequence 264, App
8	C	34	13.8	17213	15	US-10-311-455-1455	Sequence 1455, Ap
9		34	13.8	786431	15	US-10-412-277-3	Sequence 3, Appli
10	C	33.2	13.5	560	13	US-10-027-632-216434	Sequence 216434,
11	C	33.2	13.5	560	16	US-10-027-632-216434	Sequence 216434,
12	C	33.2	13.5	3759	13	US-10-282-122A-15979	Sequence 15979, A
13	C	32.8	13.3	338702	13	US-10-087-192-292	Sequence 292, App
14	C	32.6	13.3	435	13	US-10-085-783A-40780	Sequence 40780, A

C	15	32.6	13.3	435	16	US-10-242-535A-40780	Sequence 40780, A
	16	32.2	13.1	15674	15	US-10-311-455-335	Sequence 335, App
	17	32.2	13.1	15674	15	US-10-240-485-29	Sequence 29, Appl
C	18	32.2	13.1	3673778	15	US-10-312-841-2	Sequence 2, Appli
C	19	32	13.0	480	13	US-10-282-122A-15466	Sequence 15466, A
	20	32	13.0	941	13	US-10-424-599-100533	Sequence 100533,
C	21	32	13.0	3673778	15	US-10-312-841-1	Sequence 1, Appli
C	22	31.8	12.9	422	10	US-09-918-995-2677	Sequence 2677, Ap
	23	31.8	12.9	680	13	US-10-027-632-152465	Sequence 152465,
	24	31.8	12.9	680	16	US-10-027-632-152465	Sequence 152465,
	25	31.8	12.9	1885	17	US-10-641-643-1162	Sequence 1162, Ap
	26	31.8	12.9	4346	15	US-10-247-671-49	Sequence 49, Appl
	27	31.8	12.9	251364	15	US-10-175-523-58	Sequence 58, Appl
	28	31.8	12.9	251364	15	US-10-175-523-61	Sequence 61, Appl
	29	31.8	12.9	251364	15	US-10-175-523-79	Sequence 79, Appl
C	30	31.6	12.8	472	10	US-09-918-995-26533	Sequence 26533, A
C	31	31.6	12.8	580	13	US-10-027-632-237125	Sequence 237125,
C	32	31.6	12.8	580	13	US-10-027-632-237127	Sequence 237127,
C	33	31.6	12.8	580	16	US-10-027-632-237126	Sequence 237126,
C	34	31.6	12.8	580	16	US-10-027-632-237125	Sequence 237125,
C	35	31.6	12.8	580	16	US-10-027-632-237126	Sequence 237126,
C	36	31.6	12.8	580	16	US-10-027-632-237127	Sequence 237127,
C	37	31.4	12.8	1017	16	US-10-349-680-136	Sequence 136, App
	38	31.4	12.8	1215	15	US-10-264-237-600	Sequence 600, App
C	39	31.4	12.8	2939	13	US-10-342-887-301	Sequence 301, App
C	40	31.4	12.8	2939	13	US-10-172-118-301	Sequence 301, App
C	41	31.4	12.8	8467	11	US-10-311-455-81	Sequence 81, Appl
	42	31.4	12.8	8758	15	US-10-311-455-1092	Sequence 1092, Ap
C	43	31.2	12.7	400	9	US-09-960-352-8625	Sequence 8625, Ap
C	44	31.2	12.7	468	9	US-09-864-761-167	Sequence 167, App
C	45	31.2	12.7	530	9	US-09-604-287A-434	Sequence 434, App

ALIGNMENTS

RESULT 1
US-10-198-846-1685/c
; Sequence 1685, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1685
;   LENGTH: 861
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 305, 365, 383, 415, 424, 438, 443, 455, 459, 463, 472, 474,
;   LOCATION: 485, 488, 507, 535, 553, 557, 562, 567, 594, 619, 635, 637,
;   LOCATION: 646, 659, 663, 674, 678, 694, 714, 718, 725, 726, 728, 741,
;   OTHER INFORMATION: n = A,T,C or G 743, 745, 757, 759, 760, 781, 788, 810, 813, 822, 834
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 838, 858, 859
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1685

Query Match   15.6%;   Score 38.4;   DB 15;   Length 861;

Best Local Similarity　56.2%;　Pred. No. 0.19;
Matches　72;　Conservative　0;　Mismatches　56;　Indels　0;　Gaps　0;

```
Qy    24 CTTTTTATTTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCCTACTG   83
         ||    |||||||| ||| | | |||||| |||| ||
Db   235 CTTTCTGTTTGGTTTTTTTTCTTTTTTTCTATATATTTTTGTTGAAATTCTATTT        176

Qy    84 CTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCCAACTGCAAC   143
         |||                || ||||
Db   175 TATTTTTAATTCTCTCTCTCTCCCAGACACAATGGCACTGCTTATCTCCGAAATGGTGT   116

Qy   144 GAGAGTTT   151
         || ||||
Db   115 GATCGTCT   108
```

RESULT 2
US-10-282-122A-16553/c
; Sequence 16553, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;　APPLICANT: Wang, Liangsu
;　APPLICANT: Zamudio, Carlos
;　APPLICANT: Malone, Cheryl
;　APPLICANT: Haselbeck, Robert
;　APPLICANT: Ohlsen, Kari
;　APPLICANT: Zyskind, Judith
;　APPLICANT: Wall, Daniel
;　APPLICANT: Carr, Grant
;　APPLICANT: Yamamoto, Robert
;　APPLICANT: Forsyth, R.
;　APPLICANT: Xu, H.
;　TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;　FILE REFERENCE: ELITRA.034A
;　CURRENT APPLICATION NUMBER: US/10/282,122A
;　CURRENT FILING DATE: 2003-02-20
;　PRIOR APPLICATION NUMBER: 60/191,078
;　PRIOR FILING DATE: 2000-03-21
;　PRIOR APPLICATION NUMBER: 60/206,848
;　PRIOR FILING DATE: 2000-05-23
;　PRIOR APPLICATION NUMBER: 60/207,727
;　PRIOR FILING DATE: 2000-05-26
;　PRIOR APPLICATION NUMBER: 60/230,335
;　PRIOR FILING DATE: 2000-09-06
;　PRIOR APPLICATION NUMBER: 60/230,347
;　PRIOR FILING DATE: 2000-09-09
;　PRIOR APPLICATION NUMBER: 60/242,578
;　PRIOR FILING DATE: 2000-10-23
;　PRIOR APPLICATION NUMBER: 60/253,625
;　PRIOR FILING DATE: 2000-11-27
;　PRIOR APPLICATION NUMBER: 60/257,931
;　PRIOR FILING DATE: 2000-12-22
;　PRIOR APPLICATION NUMBER: 60/267,636
;　PRIOR FILING DATE: 2001-02-09
;　PRIOR APPLICATION NUMBER: 60/269,308
;　PRIOR FILING DATE: 2001-02-16
;　Remaining Prior Application data removed - See File Wrapper or PALM.
;　NUMBER OF SEQ ID NOS: 78614
;　SOFTWARE: PatentIn version 3.1
;　SEQ ID NO 16553
;　LENGTH: 3588
;　TYPE: DNA
;　ORGANISM: Clostridium botulinum
US-10-282-122A-16553

Query Match　14.7%;　Score 36.2;　DB 13;　Length 3588;
Best Local Similarity　56.2%;　Pred. No. 1.8;
Matches　68;　Conservative　0;　Mismatches　53;　Indels　0;　Gaps　0;

```
Qy     2 TCATTAACCTTTGCAAGGATACCTTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACA   61
         || ||   || || |          |||  | ||||  ||||| | | ||| |   | |
Db  3095 TCAATAACATTCAAAAGTTCTTTCTTTTTTGACTTTACTAAAATCTTCCTTTGATTATTCATA 3036
```

Best Local Similarity　56.2%;　Pred. No. 0.19;
Matches　72;　Conservative　0;　Mismatches　56;　Indels　0;　Gaps　0;

```
Qy    62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA  121
         |||   ||||||  |                ||| ||||||  | ||||  || || || ||
Db  3035 AATGTTATTTTTTCATTTATTTTTCATTTATTTCCTCTATATTCCTCTATTGATCCTACACATTTACTGTGCCT 2976

Qy   122 A 122
         |
Db  2975 A 2975
```

RESULT 3
US-09-879-536-826
; Sequence 826, Application US/0979536
; Patent No. US2002014298A1
; GENERAL INFORMATION:
;　APPLICANT: Endege, Wilson O.
;　APPLICANT: Steinmann, Kathleen E.
;　APPLICANT: Astle, Jon H.
;　APPLICANT: Burgess, Christopher C.
;　APPLICANT: Bushnell, Steven E.
;　APPLICANT: Carroll III, Eddie
;　APPLICANT: Catino, Theodore J.
;　APPLICANT: Derti, Adnan
;　APPLICANT: Ford, Donna M.
;　APPLICANT: Lewis, Marcia E.
;　APPLICANT: Monahan, John E.
;　APPLICANT: Schlegel, Robert
;　TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;　TITLE OF INVENTION: PRODUCTS
;　FILE REFERENCE: CCD-257 (US)
;　CURRENT APPLICATION NUMBER: US/09/879,536
;　CURRENT FILING DATE: 2001-09-21
;　PRIOR APPLICATION NUMBER: US 60/088,801
;　PRIOR FILING DATE: 1998-06-10
;　NUMBER OF SEQ ID NOS: 850
;　SOFTWARE: FastSEQ for Windows Version 3.0
;　SEQ ID NO 826
;　LENGTH: 507
;　TYPE: DNA
;　ORGANISM: Homo sapiens
;　FEATURE:
;　NAME/KEY: misc_feature
;　LOCATION: (1)...(507)
;　OTHER INFORMATION: n = A,T,C or G
US-09-879-536-826

Query Match　14.5%;　Score 35.6;　DB 9;　Length 507;
Best Local Similarity　62.2%;　Pred. No. 1.1;
Matches　56;　Conservative　0;　Mismatches　34;　Indels　0;　Gaps　0;

```
Qy    62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA  121
         |||||||||  | |||||   |||||||      |||||||||| |||| || || ||||
Db   203 GATTTTAAATATTACTACACCTCAGGGACCAAAAGAAAAAAGTTAAGCAAGCAGGGGTTCCA 262

Qy   122 ACCTCTACCCCCAACTGCAACGAACGAGAGTT 151
         ||       ||| || |||| |||| ||||
Db   263 AGTGCTCCTCCCCAACTTCAACAAGAATGT 292
```

RESULT 4
US-09-764-891-9844
; Sequence 9844, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
;　APPLICANT: Rosen et al.
;　TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;　FILE REFERENCE: PC006
;　CURRENT APPLICATION NUMBER: US/09/764,891
;　CURRENT FILING DATE: 2001-01-17
;　Prior application data removed - consult PALM or file wrapper
;　NUMBER OF SEQ ID NOS: 10231
;　SOFTWARE: PatentIn Ver. 2.0
;　SEQ ID NO 9844

; LENGTH: 24387
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9844

Query Match   14.5%;   Score 35.6; DB 10; Length 24387;
Best Local Similarity 62.2%;   Pred. No. 7.4;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy      62  GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA  121
            | ||||||  ||||||| | || ||| || |||  ||  |||   |  | |  ||  ||
Db   19467  GATTTTAAATATACTACACCTCAGGGACCAAAAGAAAAAGTTAAGCAAGCAGGGTTCCA  19526

Qy     122  ACCTCTACCCCCCAACTGCAACGAGAGTT  151
            |  |    | | ||||  || ||||||||
Db   19527  AGTGCTCCTCCCCAACTTCAACAAGAATGT  19556
```

RESULT 5
US-09-822-849A-479/c
; Sequence 479, Application US/09822849A
; Patent No. US2002004517A01
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 479
; LENGTH: 2803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-479

Query Match   14.4%;   Score 35.4; DB 9; Length 2803;
Best Local Similarity 63.5%;   Pred. No. 2.9;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy       2  TCATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTATACACA  61
            ||||| ||||  |||||||| || || ||||||||| ||||| |||
Db    2102  TCATTTTCCATCCCAGGATAGATATAATTTCTTTGATATTTATAATATATATATATA  2043

Qy      62  GATTTTAAGTTTACTCCTACTGCTG  86
            ||  |||  || |||||||||
Db    2042  TATATTATATGTACACACCACCTG  2018
```

RESULT 6
US-10-312-354-67/c
; Sequence 67, Application US/10312354
; Publication No. US2004010930A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; JACKSON, Jennifer L.;
; APPLICANT: TANG, Y. Tom; YUE, Henry;
; APPLICANT: ELLIOTT, Vicki S.; TRIBOULEY, Catherine M.;
; APPLICANT: LEE, Ernestine A.; RAMKUMAR, Jayalaxmi;
; APPLICANT: LAL, Preeti G.; XU, Yuming;
; APPLICANT: WARREN, Bridget A.; HAFALIA, April J. A.;
; APPLICANT: BAUGHN, Mariah R.; AZIMZAI, Yalda;
; APPLICANT: BATRA, Sajeev; BURFORD, Neil;
; APPLICANT: YAO, Monique G.; NGUYEN, Danniel B.;
; APPLICANT: LU, Dyung Aina M.; CHAWLA, Narinder K.;
; APPLICANT: GANDHI, Ameena R.; AU-YOUNG, Janice;
; APPLICANT: ARVIZU, Chandra S.
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PI-0133 USN
; CURRENT APPLICATION NUMBER: US/10/312,354
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 01/19862
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/212,890
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/213,466
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,601
; PRIOR FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: US 60/222,372
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/231,435
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/232,889
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PERL Program
; SEQ ID NO 67
; LENGTH: 4480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID NO: 1623474CB1
US-10-312-354-67

Query Match   14.4%;   Score 35.4; DB 17; Length 4480;
Best Local Similarity 63.5%;   Pred. No. 3.7;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy       2  TCATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTATACACA  61
            ||||| ||||  |||||||| || || ||||||||| ||||| |||
Db    3771  TCATTTTCCATCCCAGGATAGATATAATTTCTTTGATATTTATAATATATATATATA  3712

Qy      62  GATTTTAAGTTTACTCCTACTGCTG  86
            ||  |||  || |||||||||
Db    3711  TATATTATATGTACACACCACCTG  3687
```

RESULT 7
US-10-252-157-264
; Sequence 264, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 264
; LENGTH: 2264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 206310.2
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 481-537, 1472, 1846-1848
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-264

```
Query Match
Best Local Similarity    14.1%;    Score 34.8;  DB 15;  Length 2264;
Matches   69;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    25 TTTTTATTTTCTTTAAGATTCCTGTTGTTATACACAGATTTAAGTTTACTCCTACTGC 84
         ||||||||||||||||  ||||||| | ||||||||||||||||||   ||||||||||
Db  1452 TTCTGTTTCGTTTTTTTTTTNCTTTTTTTTCTATATATATTTTTTGTTGAATTCTATTT 1511

Qy    85 TGACCCAAGTGAAATTCCTTCCTCCAGTCACAGTGTCAACCTCTACCCCCCAACTGCAACG 144
         |||||||                |||||||||||||||||||||||||||||||||||||
Db  1512 ATTTTTAATTCTCTCTCTTCTCCTCCAGACACAATGGCACTGCTTATCTCCGAAATGGTGTG 1571

Qy   145 AGAGTTT 151
         |||||||
Db  1572 ATCGTCT 1578

RESULT 8
US-10-311-455-1455/c
; Sequence 1455, Application US/10311455
; Publication No. US2003014360GA1
; GENERAL INFORMATION:
;   APPLICANT: OLEK, Alexander
;   APPLICANT: PIEPENBROCK, Christian
;   APPLICANT: BERLIN, Kurt
;   TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;   TITLE OF INVENTION: Cytosine methylation
;   FILE REFERENCE: 5013.1014
;   CURRENT APPLICATION NUMBER: US/10/311,455
;   CURRENT FILING DATE:  2002-12-16
;   PRIOR APPLICATION NUMBER: DE 10032529.7
;   PRIOR FILING DATE:  2001-07-02
;   PRIOR APPLICATION NUMBER: PCT/EP01/07537
;   PRIOR FILING DATE:  2000-06-30
;   PRIOR APPLICATION NUMBER: DE 10043826.1
;   PRIOR FILING DATE:  2000-09-01
;   NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1455
;   LENGTH: 17213
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1455

Query Match
Best Local Similarity    13.8%;    Score 34;  DB 15;  Length 17213;
Matches   76;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     2 TCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACA 61
         ||||| |||||||||||||||||  |||||||||||||||   ||||||||||||
Db  7030 TAACAAATCTATCCTAACATCACTTTTTTTTTTTCCTATAACAATCACCAATAACTAAACAAT 6971

Qy    62 GATTTTTAAGTTTACTCCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA 121
         |||||     |||||||||||||||||||||||||||||   ||
Db  6970 ATTCACAATTAATTTTATTCAACTAATCAAATCTTAATTTTTTTTTAAAATATCACTATTA 6911

Qy   122 ACCTCTACCCCCAACTGCAACGAGA 147
         ||||||||||||
Db  6910 AATTATATCTTCTCAACTTCTATTAAA 6885

RESULT 9
US-10-412-277-3/c
; Sequence 3, Application US/10412277
; Publication No. US2003017579IA1
; GENERAL INFORMATION:
;   APPLICANT: GUEGLER, Karl et al
;   TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001067DIV
;   CURRENT APPLICATION NUMBER: US/10/412,277
;
```

```
;   CURRENT FILING DATE:  2003-04-14
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 786431
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(786431)
;   OTHER INFORMATION: n = A,T,C or G
US-10-412-277-3

Query Match
Best Local Similarity    13.8%;    Score 34;  DB 15;  Length 786431;
Matches   43;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy    16 AAGGATAACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTT 73
         |||||| |||||| | |||||  | ||||||||||||||  | |||||| ||||||||
Db 691859 ATGGATACATTTTTATAAAATTTTATACTTTAAAACACACACTTAAGATT 691802

RESULT 10
US-10-027-632-216434/c
; Sequence 216434, Application US/10027632
; Publication No. US2002019837IA1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION:  Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE:  2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE:  2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE:  2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE:  2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE:  2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE:  1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE:  1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE:  1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 216434
;   LENGTH: 560
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-216434

Query Match
Best Local Similarity    13.5%;    Score 33.2;  DB 13;  Length 560;
Matches   47;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy     7 AACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTT 66
         |||||||||||||   ||||||||||   ||||| |||| ||| ||||||||||| |||
Db   389 AAACTTTTCTTTGATATACTTTTATTATTTTCCTTTATTATACTTTTTGTTTTATTATTACTTT 330

Qy    67 TAAGTTTACT 76
         ||||||||
Db   329 CTAGTTTACT 320

RESULT 11
US-10-027-632-216434/c
; Sequence 216434, Application US/10027632
; Publication No. US2003020407SA9
; GENERAL INFORMATION:
;
```

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 216434
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-216434

Query Match   13.5%;   Score 33.2; DB 16; Length 560;
Best Local Similarity   67.1%;   Pred. No. 6.3;
Matches   47;   Conservative   0;   Mismatches   23;   Indels   0;   Gaps   0;

Qy   7 AACCTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCGTGTTGTTTATACACAGATTT 66
          ||||| ||| ||||||| ||||||||| | |||||||||||||||
Db 389 AAACTTTTCTTTGATATACTTTTATTTCCTTTATTATACTTTTGTTTATTATTTACTTT 330

Qy   67 TAAGTTTACT 76
          |||||||||
Db 329 CTAGTTTACT 320

RESULT 12
US-10-282-122A-15979/c
; Sequence 15979, Application US/10282122A
; Publication No. US2004029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15979
; LENGTH: 3759
; TYPE: DNA
; ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15979

Query Match   13.5%;   Score 33.2; DB 13; Length 3759;
Best Local Similarity   54.0%;   Pred. No. 16;
Matches   68;   Conservative   0;   Mismatches   58;   Indels   0;   Gaps   0;

Qy   20 ATACCTTTTATTTCTTTAAGATCCTGTTGTTTATACACAGATTTAAGTTTACTCCT 79
          || ||||| || |||||| || ||| ||||||||||||| || | ||||
Db 3014 ATTTCTTTAGTCTTTTATAAAATTGCGTTAGTTTTTCATTAATGTTTTTTCTTCTTCT 2955

Qy   80 ACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCAACCTTCTACCCCCAACTG 139
          | | || | | ||||| | | | |||| ||| ||||| | |||||||
Db 2954 AATAATAACTGATTAAAACTTAAAACATCATTTCTACTTAAAACATTTACTCGCCAACTA 2895

Qy 140 CAACGA 145
          | ||
Db 2894 GAAGGA 2889

RESULT 13
US-10-087-192-292/c
; Sequence 292, Application US/10087192
; Publication No.US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 52945200012
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 292
; LENGTH: 338702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(338702)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-292

Query Match   13.3%;   Score 32.8; DB 13; Length 338702;
Best Local Similarity   50.6%;   Pred. No. 2.1e+02;
Matches   79;   Conservative   0;   Mismatches   77;   Indels   0;   Gaps   0;

Qy   33 TTCTTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTTACTCCTACTGCTGACCCAA 92
          | ||| ||| || |||| || |||| | |||||| || | | | ||||||||||
Db 212965 TGCTTACCATTTATGTTGCCTTGACATCCACTTTAAAATATAGGTTAACTTTCTCAGAA 212906

Qy   93 GTGAAATTCCTTCTCCAGTCACAGTGTCAACCTTCACCCCCAACTGCAACGACGAGAGTTTT 152
          |||| |||||| | |||| |||||||||||||

Db   212905 TCAGGGCTCAGTCACCCTGACAGTTTCAAATTCCATTCCCCACCTGAATGGCTCCCACT 212846
                     | |||| |||||||| || |||||||| |||
Qy   153 GAGGGCCATCAATCACACCGAGAAGTCACAGCCCCT 188

Db   212845 GGTGCAGGAAATAAAACTGAAGAGACCCAAGTCCT 212810

RESULT 14
US-10-085-783A-40780/c
; Sequence 40780, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
;   APPLICANT: Chondrogene Inc.
;   APPLICANT: Liew, C.C.
;   TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
;   FILE REFERENCE: 4231/2002
;   CURRENT APPLICATION NUMBER: US/10/085,783A
;   CURRENT FILING DATE: 2002-02-28
;   PRIOR APPLICATION NUMBER: US 60/305,340
;   PRIOR FILING DATE: 2001-07-13
;   PRIOR APPLICATION NUMBER: US 60/275,017
;   PRIOR FILING DATE: 2001-03-12
;   PRIOR APPLICATION NUMBER: US 60/271,955
;   PRIOR FILING DATE: 2001-02-28
;   NUMBER OF SEQ ID NOS: 58994
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40780
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24)..(24)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-40780

Query Match   13.3%; Score 32.6; DB 13; Length 435;
Best Local Similarity   61.7%; Pred. No. 8.6;
Matches   50; Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   11 TTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTGTTATACACAGATTTTAAG 70
           |||||| | ||| |||||||||||| ||||| | ||| ||| |||
Db   93 TTTGTATGAAAATCACTTTATTTTCTTTAAAGTTACCATTCACTATAATATCCCATTATA 34

Qy   71 TTTACTCCTACTGCTGACCCA 91
           || ||| ||| |||||| |||
Db   33 TTCACTNCTNCTGTTGATACA 13

RESULT 15
US-10-242-535A-40780/c
; Sequence 40780, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
;   APPLICANT: Chondrogene Inc.
;   APPLICANT: Liew, C.C.
;   TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
;   FILE REFERENCE: 4231/2005
;   CURRENT APPLICATION NUMBER: US/10/242,535A
;   CURRENT FILING DATE: 2002-09-12
;   PRIOR APPLICATION NUMBER: US 10/085,783
;   PRIOR FILING DATE: 2002-02-28
;   PRIOR APPLICATION NUMBER: US 60/305,340
;   PRIOR FILING DATE: 2001-07-13
;   PRIOR APPLICATION NUMBER: US 60/275,017
;   PRIOR FILING DATE: 2001-03-12
;   PRIOR APPLICATION NUMBER: US 60/271,955
;   PRIOR FILING DATE: 2001-02-28
;   NUMBER OF SEQ ID NOS: 58994
;   SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40780
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24)..(24)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (27)..(27)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-40780

Query Match   13.3%; Score 32.6; DB 16; Length 435;
Best Local Similarity   61.7%; Pred. No. 8.6;
Matches   50; Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   11 TTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTGTTATACACAGATTTTAAG 70
           |||||| | ||| |||||||||||| ||||| | ||| ||| |||
Db   93 TTTGTATGAAAATCACTTTATTTTCTTTAAAGTTACCATTCACTATAATATCCCATTATA 34

Qy   71 TTTACTCCTACTGCTGACCCA 91
           || ||| ||| |||||| |||
Db   33 TTCACTNCTNCTGTTGATACA 13

Search completed: June 10, 2004, 17:05:34
Job time : 223 secs

```
            GenCore version 5.1.6
      Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      June 10, 2004, 15:30:58 ; Search time 2535 Seconds
                                         (without alignments)
                                       2897.868 Million cell updates/sec

Title:           US-09-674-593-10
Perfect score:   246
Sequence:        1 gtcattaaccttgcaagga........ttcatatcaaccccacacta 246

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2751328 9 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:  55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  EST:*
        1: em_estba:*
        2: em_esthum:*
        3: em_estin:*
        4: em_estmu:*
        5: em_estov:*
        6: em_estmus:*
        7: em_estpl:*
        8: em_estro:*
        9: em_htc:*
       10: gb_est2:*
       11: gb_htc:*
       12: gb_est3:*
       13: gb_est4:*
       14: gb_est5:*
       15: em_estfun:*
       16: em_estom:*
       17: em_gss_hum:*
       18: em_gss_inv:*
       19: em_gss_pln:*
       20: em_gss_vrt:*
       21: em_gss_fun:*
       22: em_gss_mam:*
       23: em_gss_mus:*
       24: em_gss_pro:*
       25: em_gss_rod:*
       26: em_gss_phg:*
       27: em_gss_vrl:*
       28: gb_gss1:*
       29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Match	Query Match	Length	DB ID	Description
	1	246	100.0	246	9	AA863443	AA863443 oh05e10.s
c	2	244.4	99.3	486	12	BM833169	BM833169 K-EST0107
	3	242	98.4	608	13	BU689416	BU689416 UI-CF-EC1
c	4	212.4	86.3	845	13	BQ948660	BQ948660 AGENCOURT
c	5	174.6	71.0	898	13	BQ650550	BQ650550 AGENCOURT
c	6	99.4	40.4	963	13	BQ650598	BQ650598 AGENCOURT
c	7	40.2	16.3	502	14	CB049560	CB049560 NISC_gj11
c	8	38	15.4	477	12	BI818248	BI818248 603032467
c	9	37.6	15.3	834	28	BH280049	BH280049 CH230-117
c	10	37.4	15.2	415	9	AI797543	AI797543 we54h01.x
c	11	37.4	15.2	433	9	AA121152	AA121152 z188e01.s
c	12	37.2	15.1	520	10	BG033754	BG033754 602302013
c	13	37.2	15.1	989	13	BX361014	BX361014 BX361014
c	14	37	15.0	1201	29	CNS001KH	AL061120 Drosophil
c	15	36.8	15.0	669	13	BW315557	BW315557 BW315557
c	16	36.6	14.9	860	28	AZ540006	AZ540006 ENTFG92TR
c	17	36.6	14.9	894	28	BH132363	BH132363 ENTON80TR
c	18	36.6	14.9	971	28	AZ687607	AZ687607 ENTHR40TF
c	19	36.4	14.8	671	29	CE637367	CE637367 tigr-gss-
c	20	36.2	14.7	945	28	BH146881	BH146881 ENTQF73TR
c	21	36.2	14.7	1201	13	BX375716	BX375716 BX375716
c	22	36	14.6	859	28	BZ154702	BZ154702 CH230-442
c	23	36	14.6	976	28	CC293061	CC293061 CH261-174
c	24	35.8	14.6	595	14	CB168554	CB168554 GKC603000
c	25	35.6	14.5	479	14	CD446415	CD446415 EL01T0201
c	26	35.6	14.5	1201	9	AL541884	AL541884 AL541884
c	27	35.4	14.4	326	12	BM984485	BM984485 UI-CF-DU1
c	28	35.4	14.4	744	9	AA056168	AA056168 zk70b12.r
c	29	35.4	14.4	928	12	BG283792	BG283792 602407504
c	30	35.4	14.2	1201	9	AL553722	AL553722 AL553722
c	31	35	14.2	846	10	BF698245	BF698245 602125586
c	32	35	14.2	1201	9	AL513871	AL513871 AL513871
c	33	34.8	14.1	319	10	AW189673	AW189673 xk93a08.x
c	34	34.8	14.1	514	28	AQ825584	AQ825584 HS_5419_A
c	35	34.8	14.1	556	29	FR0041818	AL129910 Fugu rubr
c	36	34.8	14.1	635	14	CB449540	CB449540 703747 MA
c	37	34.8	14.1	652	13	BQ389577	BQ389577 NISC_mg08
c	38	34.8	14.1	966	3	AL543767	AL543767 AL543767
c	39	34.6	14.1	1201	9	AL551035	AL551035 AL551035
c	40	34.6	14.1	422	12	BG508540	BG508540 sac74b10.
c	41	34.6	14.1	576	12	BM521549	BM521549 sak59c06.
c	42	34.6	14.1	584	12	BM521576	BM521576 sak59f06.
c	43	34.6	14.1	632	29	CE727195	CE727195 tigr-gss-
c	44	34.6	14.1	1101	29	CNS00FFA	AL070314 Drosophil
c	45	34.4	14.0	495	12	BM074816	BM074816 MEST298-A

ALIGNMENTS

```
RESULT 1
AA863443
LOCUS       AA863443                 246 bp    mRNA    linear   EST 13-MAY-1998
DEFINITION  oh05e10.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1456938 3',
            mRNA sequence.
ACCESSION   AA863443
VERSION     AA863443.1  GI:2955922
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 246)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
```

```
Insert Length: 1358   Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 208.

FEATURES          Location/Qualifiers
  source          1..246
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:1456938"
                  /lab_host="DH10B"
                  /clone_lib="NCI_CGAP_Kid3"
                  /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                  a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                  strand cDNA was primed with a Not I - oligo(dT) primer,
                  double-stranded cDNA was ligated to Eco RI adaptors
                  (Pharmacia), digested with Not I and cloned into the Not
                  I and Eco RI sites of the modified pT7T3 vector. mRNA
                  source: 2 pooled kidneys. Library went through one round
                  of normalization. Library constructed by Bento Soares and
                  M. Fatima Bonaldo.  "
ORIGIN

Query Match          100.0%;   Score 246;  DB 9;  Length 246;
Best Local Similarity 100.0%;   Pred. No. 2.9e-49;
Matches 246;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120

Qy  121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACCACCGAGAAGTCA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACCACCGAGAAGTCA 180

Qy  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240

Qy  241 ACACTA 246
          ||||||
Db  241 ACACTA 246


RESULT 2
BM833169/c
LOCUS       BM833169   486 bp   mRNA   linear   EST 06-MAR-2002
DEFINITION  K-EST0107772 S5SNU484s1 Homo sapiens cDNA clone S5SNU484s1-12-C04
            5', mRNA sequence.
ACCESSION   BM833169
VERSION     BM833169.1  GI:19189578
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 12  row: C  column: 04
```

```
High quality sequence stop: 486.

FEATURES          Location/Qualifiers
  source          1..486
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="S5SNU484s1-12-C04"
                  /sex="M"
                  /tissue_type="Stomach"
                  /cell_type="Epithelial"
                  /cell_line="SNU-484"
                  /lab_host="Top10F'"
                  /clone_lib="S5SNU484s1"
                  /note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI;
                  Site_2: NotI; The poly (A)+ RNA was decapped with tobacco
                  acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
                  including EcoRI site by treatment of T4 RNA ligase. The
                  first strand cDNA was synthesized from oligo dT-selected
                  mRNA by priming with dT-tailed vector. The dT-tailed
                  vector was adjusted to have about 60nt. The cDNA vector
                  was circularized with E. coli DNA ligase after digestion
                  of EcoRI which site is also included in vector. An RNA
                  strand converted to a DNA strand by Okayama-Berg method.
                  The obtained cDNA vectors were used for transformation of
                  competent cells E. coli Top10F' by electroporation method.
                  After analyzing and sequencing about 2,000 ~ 3,000
                  colonies in original cDNA library, the abundant cDNAs were
                  selected and amplified by PCR reaction using vector region
                  primer including T7 promotor as 5' primer and N(dT)14 as
                  3' primer. The PCR products were used as template for
                  synthesis of biotinylated single stranded RNA by in vitro
                  transcription reaction. The synthesized RNA probes were
                  hybridized with antisense single stranded cDNAs prepared
                  from original library and incubated with avidin-gel.
                  After removing DNA-RNA hybrids by centrifuge, the
                  subtracted cDNA libraries were constructed by
                  transformaion of the remaining DNA into competent cells E.
                  coli Top10F' with electroporation method."
ORIGIN

Query Match          99.3%;   Score 244.4;  DB 12;  Length 486;
Best Local Similarity 99.6%;   Pred. No. 6.7e-49;
Matches 245;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  466 GTCACTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 407

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  406 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 347

Qy  121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACCACCGAGAAGTCA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  346 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACCACCGAGAAGTCA 287

Qy  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  286 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 227

Qy  241 ACACTA 246
          ||||||
Db  226 ACACTA 221


RESULT 3
BU689416
LOCUS       BU689416   608 bp   mRNA   linear   EST 07-OCT-2002
DEFINITION  UI-CF-EC1-adw-l-20-0-UI.s1 UI-CF-EC1 Homo sapiens cDNA clone
            UI-CF-EC1-adw-l-20-0-UI 3', mRNA sequence.
ACCESSION   BU689416
VERSION     BU689416.1  GI:23547148
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 608)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..608
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-EC1-adw-1-20-0-UI"
                     /tissue_type="Lung"
                     /dev_stage="Adult and Fetal"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-EC1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-EC1 is a normalized cDNA library containing the
                     following tissue(s): Normal lung from adult and from fetal
                     day 64, day 87, week 19 and week 42. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     AAGTGCTTAC.
                     TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
                     and 380-383
                     TAG_LIB=UI-CF-EC1
                     TAG_SEQ=AAGTGCTTAC"
ORIGIN

Query Match      98.4%;  Score 242;  DB 13;  Length 608;
Best Local Similarity  100.0%;  Pred. No. 2.5e-48;
Matches  242;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      5 TTAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGAT  64
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 TTAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGAT  76

Qy     65 TTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGTCACAGTGTCAACC 124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 TTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGTCACAGTGTCAACC 136

Qy    125 TCTACCCCCAACTGCAACGAGAGTTTTGAGGGGCATCAATCACACCGAGAAGTCACAGC 184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 TCTACCCCCAACTGCAACGAGAGTTTTGAGGGGCATCAATCACACCGAGAAGTCACAGC 196
```

```
Qy    185 CCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCCACAC 244
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 CCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCCCACAC 256

Qy    245 TA 246
          ||
Db    257 TA 258
```

```
RESULT 4
BQ948660/c
LOCUS       BQ948660              845 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8784237 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6376418
            5', mRNA sequence.
ACCESSION   BQ948660
VERSION     BQ948660.1  GI:22364138
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 845)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2558 row: j  column: 03
            High quality sequence stop: 534.
FEATURES             Location/Qualifiers
     source          1..845
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6376418"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
ORIGIN

Query Match      86.3%;  Score 212.4;  DB 13;  Length 845;
Best Local Similarity  96.2%;  Pred. No. 3.5e-41;
Matches  227;  Conservative   0;  Mismatches   8;  Indels   1;  Gaps   1;

Qy     11 TTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAG 70
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||| |||||
Db    688 TTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGA-TNTAAG 630

Qy     71 TTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGTCACAGTGTCAACCTCTACC 130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 TTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGTCACAGTGTCAACCTCTACC 570

Qy    131 CCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCACAGCCCCTCA 190
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    569 CCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCACAGCCCCTCA 510

Qy    191 ACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCCCACACTA 246
```

Db   509 ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
            ACCACTGAGGTGTGGGGGGGTAGGGATCCATNTCTTCATATCAACCCCACACTA 454

RESULT 5
BQ650550/c
LOCUS   BQ650550   AGENCOURT_8350342 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6286493
            898 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION 5', mRNA sequence.
ACCESSION   BQ650550
VERSION   BQ650550.1 GI:21774722
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 898)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2487 row: g column: 06
            High quality sequence stop: 667.
FEATURES   Location/Qualifiers
     source   1..898
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6286493"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN
Query Match   71.0%; Score 174.6; DB 13; Length 898;
Best Local Similarity   94.7%; Pred. No. 4.7e-32;
Matches 213; Conservative   0; Mismatches   9; Indels   3; Gaps   3;

Qy   25 TTTTTATTTCTTTAAGATTCCTGTT-GTTTATACACAGATTTAA-GTTTACTCCTACT 82
          ||||||||||||||||||||||||| |||||||||||||||||| |||||||||||||
Db   771 TTTTTATTTCTTTAAGATTCCCGTTGGTTTATACACAGATTTTAAGGTTTACTCCTACG 712

Qy   83 GCTGACCCAAGTGAAATTCCTTCCCAGTCACAGTGTCAACCTCTA-CCCCCCAACTGCA 141
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   711 GCTGACCCAAGTGAAATTCTTTCCCAGTCACAGTGTCACAGGGTCAACCTCTACCCCCCCAACTGCA 652

Qy   142 ACGAGAGTTTTGAGGGGCATCAATCACACCGAGAAGTCACAGCCCCTCAACCACTGAGGT 201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 ACGAGAGTTTTGAGGGGCATCAATCACACCGAGAAGTCACAGCCCCCTCAACCACTGAGGT 592

Qy   202 GTGGGGGGGTAGGGATCCATCTGCATTTCTTCATATCAACCCCCACACTA 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 GTGGGGGGGTAGGGATCCATCTGCATTTCTTCATATCAACCCCCACACTA 547

RESULT 6
BQ650598/c
LOCUS   BQ650598   AGENCOURT_8207577 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6283187
            963 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION 5', mRNA sequence.
ACCESSION   BQ650598
VERSION   BQ650598.1 GI:21774770
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 963)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CGAP (Stanford)
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2478 row: m column: 12
            High quality sequence stop: 618.
FEATURES   Location/Qualifiers
     source   1..963
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6283187"
                     /tissue_type="hepatocellular carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_100"
                     /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
ORIGIN
Query Match   40.4%; Score 99.4; DB 13; Length 963;
Best Local Similarity   87.2%; Pred. No. 6.9e-14;
Matches 109; Conservative   0; Mismatches 16; Indels   0; Gaps   0;

Qy   122 ACCTCTACCCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCAC 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AACTGTCCCGCCAATGCAAGGAGAGTTTGAGGGGCATCGAATCACACCGAGAAGTCCC 602

Qy   182 AGCCCCTCAACCACTGAGGTGTGGGGGGTAGGATCTGCATTTCTTCATATCAACCCCA 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCCA 542

Qy   242 CACTA 246
          |||||
Db   541 CACTA 537

RESULT 7
CB049560/c
LOCUS   CB049560   NISC_gj11f04.y1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3271446
            502 bp   mRNA   linear   EST 17-JAN-2003
DEFINITION 5', mRNA sequence.
ACCESSION   CB049560
VERSION   CB049560.1 GI:27787847
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1432 row: e column: 18
High quality sequence stop: 408.

FEATURES   Location/Qualifiers
source   1..477
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:5173625"
                  /lab_host="DH10B"
                  /clone_lib="NIH_MGC_115"
                  /note="Organ: pooled brain, lung, testis; Vector:
                  pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                  source anonymous pool of 6 male brains, age range 23-27; 1
                  male lung, age 27; and 1 male testis, age 69. Library is
                  oligo-dT primed and directionally cloned (EcoRV site is
                  destroyed upon cloning). Average insert size 1.8 kb,
                  insert size range 1-3 kb. Library is normalized and
                  enriched for full-length clones and was constructed by C.
                  Gruber (Invitrogen). Research Genetics tracking code
                  021. Note: this is a NIH_MGC Library."

ORIGIN

Query Match   15.4%;   Score 38; DB 12; Length 477;
Best Local Similarity 55.2%; Pred. No. 50;
Matches   74; Conservative   0; Mismatches   60; Indels   0; Gaps   0;

Qy   25 TTTTTATTTTCTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTTACTCCTACTGC 84
         ||||||| |||||| |||||||||||||||||||||||| ||||||||||| ||||||
Db   381 TTTTTTTTTTTTTTTTTAAGTTTTTTTTTTTTTTCAGTTCAGCTTGGTGGTTTCCTCTTTATTG 322

Qy   85 TGACCCAAGTGAAATTCCTTCTCCAGTCCACAGTGTCAACCTTCACCCCCAACTGCAACG 144
         |||||||||||| ||||||| ||||||||||| |||||| |||||||| |||||
Db   321 ATGTGCCTCCTACCTTCCCCCCACAATTTCAGTCCCTTCCACTACCCCCAAAAAGAAGGT 262

Qy   145 AGAGTTTTGAGGGG 158
         |||| | |||| ||
Db   261 AGTGAAAGGAAGGG 248

RESULT 9
BH280049/c
LOCUS   BH280049   834 bp   DNA   linear   GSS 30-NOV-2001
DEFINITION   CH230-117E16.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
             CH230-117E16, genomic survey sequence.
ACCESSION   BH280049
VERSION   BH280049.1 GI:17192451
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 834)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
             Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
             Riggs,F., de Jong,P. and Fraser,C.M.
TITLE   Rat BAC End Sequences from Library CHORI-230 EcoRI segment
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: CH230-117E16.TV
             Contact: Shaying Zhao
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0200
             Fax: 301 838 0208

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             info@image.llnl.gov
             Plate: LLAM8008 row: L column: 7
             Seq primer: M13Rp1 reverse primer (ABI).

FEATURES   Location/Qualifiers
source   1..502
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3271446"
                  /sex="male"
                  /dev_stage="adult"
                  /lab_host="DH10B"
                  /clone_lib="NCI_CGAP_Pr28"
                  /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
                  with a modified polylinker; Plasmid DNA from the
                  normalized library NCI_CGAP_Pr22 was prepared, and ss
                  circles were made in vitro. Following HAP purification,
                  this DNA was used as tracer in a subtractive hybridization
                  reaction. The driver was PCR-amplified cDNAs from a pool
                  of 5,000 clones made from the same library (cloneIDs
                  985508-986759, 1101192-1101959, and 1217928-1220615).
                  Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match   16.3%;   Score 40.2; DB 14; Length 502;
Best Local Similarity 57.6%; Pred. No. 15;
Matches   72; Conservative   0; Mismatches   53; Indels   0; Gaps   0;

Qy   25 TTTTTATTTTCTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTTACTCCTACTGC 84
         ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   437 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA 378

Qy   85 TGACCCAAGTGAAATTCCTTCTCCAGTCCACAGTGTCAACCTTCACCCCCAACTGCAACG 144
         ||||||||| || |||||||||||||||||||||||||||| ||||||
Db   377 TAAGACTTTTCAGATACTTTTCCCAGTGCATAGTGTGAATCTTCCCAAAAGCAAGTGAAACA 318

Qy   145 AGAGT 149
         |||||
Db   317 TGAAT 313

RESULT 8
BI818248/c
LOCUS   BI818248   477 bp   mRNA   linear   EST 04-OCT-2001
DEFINITION   603032467F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5173625 5',
             mRNA sequence.
ACCESSION   BI818248
VERSION   BI818248.1 GI:15928686
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 477)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.

Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 117 row: E column: 16
Seq primer: SP6
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1..834
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-117E16"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

Query Match             15.3%;   Score 37.6;  DB 28;  Length 834;
Best Local Similarity   48.6%;   Pred. No. 60;
Matches  103;  Conservative    0;  Mismatches  109;  Indels  0;  Gaps  0;

Qy      9 CCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTA 68
            |||  |||  ||  | |||||||||||||| |      |||||||||||||||||
Db    367 CCTGTGCTAAGAAGTGTTTGATTTTGTTTTTTTTATTTCAGTCAACTAAAAAGCTTTCTGC 308

Qy     69 AGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGCACAGTGTCAACCTCTA 128
            |||||||||||||||||||||||||||||||||||||||||
Db    307 CATTCCTGCCCAACTTCTAAACCTAAGTGGTCTTCCTTCTCAAATAAATGGTTCTAAAGCTG 248

Qy    129 CCCCCCAACTGCAACGAGAGAGTTTTGAGGGGGCATCAATCACACCGAGAAGTCACAGCCCCT 188
            |||||||||||||||||||||
Db    247 TTTCTTGAGTCCAGATAGACTTTATTATGAGAACACCATTAGATTTTAAAGCCCCCTTCGCCT 188

Qy    189 CAACCACTGAGGTGTGTGGGGGTAGGGATCTG 220
            ||||||||||||||
Db    187 GTACAAGGTGGGGGCCGAGTGTAACAGTTAGCTG 156
```

RESULT 10
AI797543/c
LOCUS   AI797543   415 bp   mRNA   linear   EST 18-DEC-1999
DEFINITION we54h01.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2344945 3'
            similar to contains MER22.b2 MER22 repetitive element ;, mRNA
            sequence.
ACCESSION   AI797543
VERSION   AI797543.1 GI:5363015
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 415)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
            M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1718   Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 414.

```
FEATURES             Location/Qualifiers
     source          1..415
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2344945"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Co3"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I, Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization. "
ORIGIN

Query Match             15.2%;   Score 37.4;  DB 9;  Length 415;
Best Local Similarity   55.9%;   Pred. No. 71;
Matches   71;  Conservative    0;  Mismatches   56;  Indels  0;  Gaps  0;

Qy     25 TTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCCTACTGC 84
            |||||  |||||| | ||||||||||||||  |    ||||||||||||| |||||||||||
Db    250 TTTCTGTTTCGTTTTTTTTTTTCTTTTTTTCTATATATATTTTTGTTGAATTCTATTTT 191

Qy     85 TGACCCAAGTGAAATTCCTTCTCCAGTGCACAGTGTCAACCTCTCCACCCCCCACTGCAACG 144
            |||||||||||||||||||||||||
Db    190 ATTTTTAATTCTCTCTCTCTCCTCCAGACACAATGGCACTGCTTATCTCCGAAATGGTGTG 131

Qy    145 AGAGTTT 151
            |||| ||
Db    130 ATCGTCT 124
```

RESULT 11
AA121152/c
LOCUS   AA121152   433 bp   mRNA   linear   EST 19-MAY-1997
DEFINITION z188e01.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
            IMAGE:511704 3', mRNA sequence.
ACCESSION   AA121152
VERSION   AA121152.1 GI:1678695
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 433)
AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
            Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
            Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
            Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J.,
            Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
            and Marra,M.
TITLE   Generation and analysis of 280,000 human expressed sequence tags
JOURNAL   Genome Res. 6 (9), 807-828 (1996)
MEDLINE   97044478
PUBMED   8889549
COMMENT   Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

```
FEATURES
        source        Insert Length: 747    Std Error: 0.00
                      Seq primer: -40M13 fwd. from Amersham
                      High quality sequence stop: 371.
                      1..433
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="GDB:3844199"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:511704"
                      /tissue_type="tumor"
                      /cell_line="T84 carcinoma cell line"
                      /lab_host="SOLR cells (kanamycin resistant)"
                      /clone_lib="Stratagene colon (#937204)"
                      /note="Organ: colon; Vector: pBluescript SK-; Site 1:
                      EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
                      Oligo dT. T-84 colonic epithelial cell line. Average
                      insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor
                      sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTTTT 3'"
ORIGIN

Query Match        15.2%;   Score 37.4;  DB 9;   Length 433;
Best Local Similarity 55.9%;   Pred. No. 70;
Matches  71;  Conservative   0;  Mismatches   56;  Indels   0;  Gaps   0;

Qy      25  TTTTTATTTTCTTAAGATTCCTGTTGTTATACACAGATTTAAGTTTACTCCTACTGC  84
            |||| ||  |||||| |||| ||| ||| |||  ||||||||||||||||||||||||
Db     244  TTTCTGTTTCGTTTTTTTTTTTTCTTTTTTTTCTATATATTTTTGTTGAATTCTATTTT 185

Qy      85  TGACCCAAGTGAAATTCCTTCTCCCAGTCACAGTGCAACCTCTACCCCCAACTGCAACG 144
            |||||||||||||||||||||||||  ||||| ||||||||||||||||||||||||||
Db     184  ATTTTTAATTCTCTCTCTCCCCAGACACAATGGCACTGCTGCTTATCTCCGAAATGGTGTG 125

Qy     145  AGAGTTT 151
            |||||||
Db     124  ATCGTCT 118

RESULT 12
BG033754/c
LOCUS        BG033754             520 bp    mRNA      linear   EST 24-JAN-2001
DEFINITION   602302013F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403426 5',
             mRNA sequence.
ACCESSION    BG033754
VERSION      BG033754.1    GI:12426208
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 520)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10113  row: j  column: 03
             High quality sequence stop: 223.
FEATURES
        source        1..520
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4403426"
                      /tissue_type="mammary adenocarcinoma, cell line"
```

```
                      /lab_host="DH10B (phage-resistant)"
                      /clone_lib="NIH_MGC_87"
                      /note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
                      Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 1.383 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
ORIGIN

Query Match        15.1%;   Score 37.2;  DB 10;   Length 520;
Best Local Similarity 53.4%;   Pred. No. 78;
Matches  78;  Conservative   0;  Mismatches   68;  Indels   0;  Gaps   0;

Qy      25  TTTTTATTTTCTTAAGATTCCTGTTGTTATACACAGATTTAAGTTTACTCCTACTGC  84
            |||||||||||||||| |||| ||||||||||||||| |||||||||||||||||||||
Db     192  TTTTTTTTTTTTTTTTTTTTTTGCATATATACAACTACTTATTTTTATTGACTTT 133

Qy      85  TGACCCAAGTGAAATTCCTTCTCCCAGTCACAGTGCAACCTCTACCCCCAACTGCAACG 144
            |||||||||||||||||||  |||||||  ||||||| ||||||  |||||||||||||
Db     132  ATATAAATAGTGAAATCCCTTTAAGCAACCTAGGTATACAGTGGTGTCCAACTTTCAGG 73

Qy     145  AGAGTTTTGAGGGGCATCAATCACAC 170
            |||||||||||||  |||||
Db      72  GGATTTTCGGGGGGAGTAGAAACAC 47

RESULT 13
BX361014/c
LOCUS        BX361014             989 bp    mRNA      linear   EST 05-MAY-2003
DEFINITION   BX361014 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
             clone CS0DI078YP09 5-PRIME, mRNA sequence.
ACCESSION    BX361014
VERSION      BX361014.1    GI:30382486
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 989)
AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. This sequence belongs to sequence cluster 10967.f For
             more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CS0DI078CH05QP1&cluster=10967.f. Contact :
             Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CS0DI078CH05QP1.
FEATURES
        source        1..989
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DI078YP09"
                      /tissue_type="PLACENTA COT 25-NORMALIZED"
                      /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                      /note="1st strand cDNA was primed with a NotI-oligo(dT)
                      primer. Five prime end enriched, double-strand cDNA was
                      digested with Not I and cloned into the Not I and EcoR V
                      sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match        15.1%;   Score 37.2;  DB 13;   Length 989;
Best Local Similarity 30.1%;   Pred. No. 74;
Matches  68;  Conservative   37;  Mismatches  121;  Indels   0;  Gaps   0;

Qy       3  CATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTTATACACAG  62
```

Qy   121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCA 167
              :: ||| | :| | :||| || ||||| |||
Db   266 WWATGWGCTTCCCCATTCGTWAAATCAGTWTTWATGWAGTTTCWTCA 220

RESULT 15
BW315557
LOCUS   BW315557   669 bp   mRNA   linear   EST 11-NOV-2002
DEFINITION   BW315557 Nori Satoh unpublished cDNA library, heart Ciona
             intestinalis cDNA clone ciht029c24 5', mRNA sequence.
ACCESSION   BW315557
VERSION   BW315557.1 GI:24896241
KEYWORDS   EST.
SOURCE   Ciona intestinalis
ORGANISM   Ciona intestinalis
             Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
             Phlebobranchia; Cionidae; Ciona.
REFERENCE   1 (bases 1 to 669)
AUTHORS   Satou,Y., Shin-i,T., Kohara,Y. and Satoh,N.
TITLE   Expressed genes in Ciona intestinalis (2002c)
JOURNAL   Unpublished (2002)
COMMENT   Contact: Nori Satoh
             Department of Zoology
             Kyoto University
             Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
             Tel: 81-75-753-4081
             Fax: 81-75-705-1113
             Email: satoh@ascidian.zool.kyoto-u.ac.jp.
FEATURES   Location/Qualifiers
     source   1..669
                     /organism="Ciona intestinalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:7719"
                     /clone="ciht029c24"
                     /tissue_type="heart"
                     /clone_lib="Nori Satoh unpublished cDNA library, heart"
ORIGIN

Query Match   15.0%;   Score 36.8; DB 13; Length 669;
Best Local Similarity 61.5%; Pred. No. 95;
Matches   59; Conservative   0; Mismatches 37; Indels   0; Gaps   0;

Qy   3 CATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAG 62
            ||||||| |||| ||||| ||| || :|||||||||||| || ||||:|||||||||
Db   547 CATTATGGGTGCCTGGGTGTCATTTTATAACTTTTGAAATTTTATAATTTTGAAACAGGG 606

Qy   63 ATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAA 98
            ||:|||||| ||||||||||:|||||||||
Db   607 GTTTATGTTTATTTCTAGTTATAAACAAAGGTAAA 642

Search completed: June 10, 2004, 17:53:59
Job time : 2543 secs

Db   775 CYTTATATATYNYCCMMAMTTTTTTTMACAAWTTYTWHACMCCCCTYYTYTYTTTTYTCAC 716
              ||| ||| | :|| | :|:: || || ||||
Qy   63 ATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCAA 122
              :: |||| | :|: | || :|| :|::| ::|||
Db   715 MAAAWWAWTTTWTTTHHAMMCMCYTCCCYATMMCMCYYYTHTMMAAMMCYMWTTCMC 656

Qy   123 CCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCACA 182
              :|: |||| ||| :   :
Db   655 MCCCYCWTHCCCCSSCSNNNNNNNNSSSNNNNNNNNSSNGNNCSNNNNNNNNNNNNN 596

Qy   183 GCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTT 228
              ||||||||||
Db   595 NGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNTTTTTT 550

RESULT 14
CNS001KH/c
LOCUS   CNS001KH   1201 bp   DNA   linear   GSS 03-JUN-1999
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
             BACR04C08 of RPCI-98 library from Drosophila melanogaster (fruit
             fly), genomic survey sequence.
ACCESSION   AL061120
VERSION   AL061120.1 GI:4939583
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
             Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
             Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
             Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1201)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
             - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org The BDGP Drosophila
             melanogaster BAC library was prepared by Kazutoyo Osoegawa and
             Aaron Mammoser in Pieter de Jong's laboratory in the Department of
             Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
             NY. The library is named RPCI-98 and was constructed by partial
             EcoRI digestion of Drosophila DNA provided by the BDGP from the
             isogenic strain y2; cn bw sp, the same strain used for the BDGP's
             P1 and EST libraries. A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
     source   1..1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR04C08"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

Query Match   15.0%;   Score 37; DB 29; Length 1201;
Best Local Similarity 38.3%; Pred. No. 81;
Matches   64; Conservative 29; Mismatches 74; Indels   0; Gaps   0;

Qy   1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
            | :||: || ||::|| :|||:|| :|||| :|:||| :|||| :: || |:|| ||
Db   386 KTTYTKWTGTTGTACWWGCYWAWGVWTTTGYTCTACWCCTGYAYTYTKWTTTTCAGCCWW 327

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
            || :|||||| |||| ||| |: | :|||||||||||:||||||||||||||:
Db   326 WTKWTWTTKWTTTCCYGCAWACWGWAAGGGAWGTWGATWACCTTCYGAXTCACCCTGCC 267

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: June 10, 2004, 15:30:58 ; Search time 1706 Seconds
                                         (without alignments)
                         6249.933 Million cell updates/sec

Title: US-09-674-593-10
Perfect score: 246
Sequence: 1 gtcattaaccttgcaagga.........ttcatatcaacccacacta 246

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
               1: gb_ba:*
               2: gb_htg:*
               3: gb_in:*
               4: gb_om:*
               5: gb_ov:*
               6: gb_pat:*
               7: gb_ph:*
               8: gb_pl:*
               9: gb_pr:*
              10: gb_ro:*
              11: gb_sts:*
              12: gb_sy:*
              13: gb_un:*
              14: gb_vi:*
              15: em_ba:*
              16: em_fun:*
              17: em_hum:*
              18: em_in:*
              19: em_mu:*
              20: em_om:*
              21: em_or:*
              22: em_ov:*
              23: em_pat:*
              24: em_ph:*
              25: em_pl:*
              26: em_ro:*
              27: em_sts:*
              28: em_un:*
              29: em_vi:*
              30: em_htg_hum:*
              31: em_htg_inv:*
              32: em_htg_other:*
              33: em_htg_mus:*
              34: em_htg_pln:*
              35: em_htg_rod:*
              36: em_htg_mam:*
              37: em_htg_vrt:*
              38: em_sy:*
              39: em_htgo_hum:*
              40: em_htgo_mus:*
              41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	246	100.0	246	6	BD206172	BD206172 Tumor-ass
2	244.4	99.3	1382	6	BD206166	BD206166 Tumor-ass
3	244.4	99.3	1382	9	AF181722	AF181722 Homo sapi
4	244.4	99.3	4377	9	AF181720	AF181720 Homo sapi
5	244.4	99.3	120029	2	HSJ282H10	AL132672 Homo sapi
6	244.4	99.3	152966	9	AL359713	AL359713 Human DNA
7	39.8	16.2	211530	2	AC106070	AC106070 Rattus no
8	38.6	15.7	66522	2	AC122565	AC122565 Mus muscu
9	38.6	15.7	176244	10	AC129208	AC129208 Mus muscu
10	37.8	15.4	100902	5	AL929176	AL929176 Zebrafish
11	37.6	15.3	200985	10	AC121787	AC121787 Mus muscu
12	37.6	15.3	242754	2	AC129445	AC129445 Rattus no
13	37.6	15.3	212404	10	AC105857	AC105857 Rattus no
14	37.4	15.2	160699	2	AC090127	AC090127 Mus muscu
15	37.2	15.1	223974	2	AC138731	AC138731 Pongo pyg
16	37	15.0	227302	2	AC106343	AC106343 Rattus no
17	37	15.0	240977	10	AC106442	AC106442 Rattus no
18	36.8	15.0	103179	10	AL513352	AL513352 Mouse DNA
19	36.6	14.9	220936	2	BX537301	BX537301 Mouse DNA
20	36.6	14.9	216849	10	AC020256	AC020256 Drosophil
21	36.2	14.7	243290	10	AC129333	AC129333 Mus muscu
22	36.2	14.7	87065	10	AL663088	AL663088 Mouse DNA
23	36	14.6	110000	2	AC095863_05	Continuation (6 of
24	36	14.6	120330	8	AP004261	AP004261 Oryza sat
25	36	14.6	141293	2	AP004337	AP004337 Oryza sat
26	36	14.6	142446	8	AP005126	AP005126 Oryza sat
27	36	14.6	147750	10	AC122503	AC122503 Mus muscu
28	36	14.6	180230	10	AC124692	AC124692 Mus muscu
29	36	14.6	185066	10	AL844560	AL844560 Mouse DNA
30	36	14.6	218951	2	AC111418	AC111418 Rattus no
31	36	14.6	259077	2	AC105877	AC105877 Rattus no
32	36	14.6	135389	9	HSJ1174H9	AL121880 Human DNA
33	35.8	14.6	181016	9	AL512292	AL512292 Human DNA
34	35.8	14.6	254160	2	AC068273	AC068273 Homo sapi
35	35.8	14.6	507	6	AC060761	AC060761 Mus muscu
36	35.6	14.5	43661	9	BD210936	BD210936 Human gen
37	35.6	14.5	132029	9	AC004493	AC004493 Homo sapi
38	35.6	14.5	148724	2	HS79SG23	AL031000 Human DNA
39	35.6	14.5	155017	9	AC010778	AC010778 Homo sapi
40	35.6	14.5	186278	9	AC068583	AC068583 Homo sapi
41	35.6	14.5	204521	2	AC079176	AC079176 Homo sapi
42	35.6	14.5	259508	2	AC125005	AC125005 Mus muscu
43	35.6	14.5	630	11	AC112545	AC112545 Rattus no
44	35.4	14.4	1800	9	BV004534	BV004534 S208P6386
45	35.4	14.4			HSM801200	AL117662 Homo sapi

ALIGNMENTS

RESULT 1
BD206172
LOCUS   BD206172   246 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Tumor-associated antigen encoded by reverse strand of novel gene
            expressed unevenly.
ACCESSION   BD206172
VERSION   BD206172.1 GI:33015942
KEYWORDS   JP 200251400-A/7.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 246)
  AUTHORS   Eynde.B.V.D. and Falleur.T.B.
  TITLE   Tumor-associated antigen encoded by reverse strand of novel gene

Pred. No. is the number of results predicted by chance to have a

RESULT 2
BD206166

LOCUS	BD206166	1382 bp DNA linear PAT 17-JUL-2003
DEFINITION	Tumor-associated antigen encoded by reverse strand of novel gene expressed unevenly.	
ACCESSION	BD206166	
VERSION	BD206166.1 GI:33015936	
KEYWORDS	JP 200251400-A/1.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 1382)	
AUTHORS	Eynde,B.V.D. and Falleur,T.B.	
TITLE	Tumor-associated antigen encoded by reverse strand of novel gene expressed unevenly	
JOURNAL	Patent: JP 200251400-A 1 21-MAY-2002;	
	LUDWIG INSTITUTE FOR CANCER RESEARCH	
COMMENT	OS Homo sapiens (human)	
	PN JP 200251400-A/1	
	PD 21-MAY-2002	
	PF 13-MAY-1999 JP 2000548350	
	PR 13-MAY-1998 US 60/085318	
	PI BENOIT VAN DEN EYNDE,THIERRY BOON FALLEUR	
	PC C12N15/09,A61K38/00,A61K39/00,A61P35/00,C07K14/47,C07K16/18,	
	PC C12N1/15,	
	PC C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12P21/08,C12Q1/68// PC	
	A61K35/12,	
	PC C12N15/00,A61K37/02,C12N5/00	
	CC Tumor-associated antigen encoded by reverse strand of novel	
	CC gene expressed	
	CC unevenly.	
	FH Key Location/Qualifiers	
	FT source 1..246	
	FT /organism="Homo sapiens"	
	FT /mol_type="genomic DNA"	
	FT /db_xref="taxon:9606"	

FEATURES   Location/Qualifiers
     source   1..246
                     /organism='Homo sapiens (human)'.

ORIGIN

Query Match   100.0%; Score 246; DB 6; Length 246;
Best Local Similarity   100.0%; Pred. No. 1.7e-61;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 GTCATTAACCTTTGCAAGGATACCTTTTCTTTAATTTCTTTAAGATTCCTGTTGTTTATACAC 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GTCATTAACCTTTGCAAGGATACCTTTTCTTTAATTTCTTTAAGATTCCTGTTGTTTATACAC 60

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120

Qy  121 AACCTCTACCCCCACCACTGCAACGAGAGTTTGAGGGGGCATCAATCACACCGAGAAGTCA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AACCTCTACCCCCACCACTGCAACGAGAGTTTGAGGGGGCATCAATCACACCGAGAAGTCA 180

Qy  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240

Qy  241 ACACTA 246
        ||||||
Db  241 ACACTA 246
```

RESULT 3
AF181722

LOCUS	AF181722	1382 bp mRNA linear PRI 10-JAN-2000
DEFINITION	Homo sapiens RU2AS (RU2) mRNA, complete cds.	
ACCESSION	AF181722	
VERSION	AF181722.1 GI:6684531	
KEYWORDS	.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 1382)	
AUTHORS	Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L.,	
	Devuyst,O., Lorge,F., Weynants,P. and Boon,T.	
TITLE	A new antigen recognized by cytolytic T lymphocytes on a human	
	kidney tumor results from reverse strand transcription	
JOURNAL	J. Exp. Med. 190 (12), 1793-1800 (1999)	
MEDLINE	20069887	
PUBMED	10601354	
REFERENCE	2 (bases 1 to 1382)	
AUTHORS	Van den Eynde,B.J., Gaugler,B. and Pilotte,L.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,	
	Avenue Hippocrate, 74, Brussels 1200, Belgium	
FEATURES	Location/Qualifiers	
	source 1..1382	
	/organism="Homo sapiens"	
	/mol_type="mRNA"	

JOURNAL   expressed unevenly
                     Patent: JP 200251400-A 7 21-MAY-2002;
                     LUDWIG INSTITUTE FOR CANCER RESEARCH
COMMENT   OS Homo sapiens (human)
                 PN JP 200251400-A/7
                 PD 21-MAY-2002
                 PF 13-MAY-1999 JP 2000548350
                 PR 13-MAY-1998 US 60/085318
                 PI BENOIT VAN DEN EYNDE,THIERRY BOON FALLEUR
                 PC C12N15/09,A61K38/00,A61K39/00,A61P35/00,C07K14/47,C07K16/18,
                 PC C12N1/15,
                 PC C12N1/19,C12N1/21,C12N5/10,C12P21/02,C12P21/08,C12Q1/68// PC
                 A61K35/12,
                 PC C12N15/00,A61K37/02,C12N5/00
                 CC Tumor-associated antigen encoded by reverse strand of novel
                 CC gene expressed
                 CC unevenly.
                 FH Key   Location/Qualifiers
                 FT CDS   738..989.
                 FT source   1..1382
                 FT   /organism="Homo sapiens"
                 FT   /mol_type="genomic DNA"
                 FT   /db_xref="taxon:9606"

ORIGIN

Query Match   99.3%; Score 244.4; DB 6; Length 1382;
Best Local Similarity   99.6%; Pred. No. 5.3e-61;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    1 GTCATTAACCTTTGCAAGGATACCTTTTCTTTAATTTCTTTAAGATTCCTGTTGTTTATACAC 60
        |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  268 GTCACTAACCTTTGCAAGGATACCTTTTCTTTAATTTCTTTAAGATTCCTGTTGTTTATACAC 327

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  328 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 387

Qy  121 AACCTCTACCCCCACCACTGCAACGAGAGTTTGAGGGGGCATCAATCACACCGAGAAGTCA 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  388 AACCTCTACCCCCACCACTGCAACGAGAGTTTGAGGGGGCATCAATCACACCGAGAAGTCA 447

Qy  181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  448 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 507

Qy  241 ACACTA 246
        ||||||
Db  508 ACACTA 513
```

```
        gene          /db_xref="taxon:9606"
                      /chromosome="6"
                      /map="6p22.1"
                      1..1382
        CDS           /gene="RU2"
                      738..992
                      /gene="RU2"
                      /product="RU2AS"
                      /codon_start=1
                      /protein_id="AAF23613.1"
                      /db_xref="GI:6684532"
                      /translation="MDDDAAPRVEGVPVAVHKHALHDGLRQVAGPGAAAAHLPRWPPP
                      QLAASRREAPPLSQRPHRTQGAGSPPETNEKLTNPQVKEK"
        ORIGIN

        Query Match   99.3%;   Score 244.4;  DB 9;  Length 1382;
        Best Local Similarity  99.6%;  Pred. No. 5.3e-61;
        Matches 245;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

        Qy      1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db    268 GTCACTAACCTTTGCAAGGATACCTTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 327

        Qy     61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db    328 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 387

        Qy    121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 180
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db    388 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 447

        Qy    181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGAGGATCTGCATTTCTTCATATCAACCCC 240
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db    448 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGAGGATCTGCATTTCTTCATATCAACCCC 507

        Qy    241 ACACTA 246
                  ||||||
        Db    508 ACACTA 513

        RESULT 4
        AF181720/c
        LOCUS       AF181720       4377 bp   DNA      linear   PRI 10-JAN-2000
        DEFINITION  Homo sapiens RU2AS (RU2) gene, complete cds; and RU2S (RU2) gene,
                    partial cds.
        ACCESSION   AF181720
        VERSION     AF181720.1  GI:6684526
        KEYWORDS    .
        SOURCE      Homo sapiens (human)
          ORGANISM  Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
        REFERENCE   1  (bases 1 to 4377)
          AUTHORS   Van Den Eynde,B.J., Gaugler,B., Probst-Kepper,M., Michaux,L.,
                    Devuyst,O., Lorge,F., Weynants,P. and Boon,T.
          TITLE     A new antigen recognized by cytolytic T lymphocytes on a human
                    kidney tumor results from reverse strand transcription
          JOURNAL   J. Exp. Med. 190 (12), 1793-1800 (1999)
          MEDLINE   20069887
          PUBMED    10601354
        REFERENCE   2  (bases 1 to 4377)
          AUTHORS   Van den Eynde,B.J., Gaugler,B. and Pilotte,L.
          TITLE     Direct Submission
          JOURNAL   Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,
                    Avenue Hippocrate, 74, Brussels 1200, Belgium
        FEATURES            Location/Qualifiers
          source    1..4377
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /map="6p22.1"
          gene      1754..>2348
```

```
        mRNA          /gene="RU2"
                      1754..>2348
                      /gene="RU2"
                      /product="RU2S"
                      complement(<1912..>2166)
        mRNA          /product="RU2AS"
                      complement(1912..2166)
        CDS           /product="RU2AS"
                      /codon_start=1
                      /product="RU2AS"
                      /protein_id="AAF23611.1"
                      /db_xref="GI:6684528"
                      /translation="MDDDAAPRVEGVPVAVHKHALHDGLRQVAGPGAAAAHLPRWPPP
                      QLAASRREAPPLSQRPHRTQGAGSPPETNEKLTNPQVKEK"
        CDS           2056..>2348
                      /gene="RU2"
                      /codon_start=1
                      /product="RU2S"
                      /protein_id="AAF23610.1"
                      /db_xref="GI:6684527"
                      /translation="MSGGSARSSHLSQPVVKSVLVRNGDPFYAGRVVIHEKKVSSF
                      EVFLKEVTGGVQAFFGAVRNIYTPRTGHRIRKLDIQSGGNYVAGGQEAFKKL"
        ORIGIN

        Query Match   99.3%;   Score 244.4;  DB 9;  Length 4377;
        Best Local Similarity  99.6%;  Pred. No. 5.6e-61;
        Matches 245;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

        Qy      1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db   2636 GTCACTAACCTTTGCAAGGATACCTTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 2577

        Qy     61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db   2576 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 2517

        Qy    121 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCAATCACACCGAGAAGTCA 180
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db   2516 AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCAATCACACCGAGAAGTCA 2457

        Qy    181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGAGGATCTGCATTTCTTCATATCAACCCC 240
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        Db   2456 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGAGGATCTGCATTTCTTCATATCAACCCC 2397

        Qy    241 ACACTA 246
                  ||||||
        Db   2396 ACACTA 2391

        RESULT 5
        HSJ282H10
        LOCUS       HSJ282H10         120029 bp   DNA   linear   HTG 10-JUL-2001
        DEFINITION  Homo sapiens chromosome 6 clone RP1-282H10 map p22.1-22.3, 5
                    unordered pieces.
        ACCESSION   AL132672
        VERSION     AL132672.14  GI:14348905
        KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
        SOURCE      Homo sapiens (human)
          ORGANISM  Homo sapiens
                    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
        REFERENCE   1
          AUTHORS   Garner,P.
          TITLE     Direct Submission
          JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
                    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                    requests: clonerequest@sanger.ac.uk
        COMMENT     On Jun 12, 2001 this sequence version replaced gi:12331282.
                    -------------- Genome Center
                    Center: Sanger Centre
                    Center code: SC
                    Web site: http://www.sanger.ac.uk
                    Contact: humquery@sanger.ac.uk
                    -------------- Project Information
```

```
Db    21435  AACCTCTACCCCCAACTGCAACGAGAGTTTTGAGGGGCATCAATCACCGAGAAGTCA  21494
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      181   CAGCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC  240

Db    21495  CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC  21554
                |||||||||||||| |||||||||| |||||||||||||||||||||||||||||||||
Qy      241   ACACTA  246
                ||||||
Db    21555  ACACTA  21560
```

```
RESULT 6
AL359713
LOCUS       AL359713             152966 bp    DNA        linear   PRI 01-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-95P3 on chromosome 6, complete
            sequence.
ACCESSION   AL359713
VERSION     AL359713.25  GI:13938809
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152966)
  AUTHORS   Kimberley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 3, 2001 this sequence version replaced gi:13446455.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-95P3 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-95P3 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-95P3 is at 1 in this sequence. The
            true left end of clone RP1-73M23 is at 152867 in this sequence. The
            true right end of clone RP11-40B20 is at 17700 in this sequence.
FEATURES             Location/Qualifiers
     source          1..152966
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-95P3"
                     /clone_lib="RPCI-11.1"
     misc_feature    19..544
                     /note="match: GSS: Em:AZ517849"
     misc_feature    24..379
```

```
Center project name: dJ282H10
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 0% of reads
Sequencing vector: plasmid; L08752; 99% of reads
Chemistry: Dye-terminator ABI; 1% of reads
Chemistry: Dye-terminator Big Dye; 98% of reads
Consensus quality: 118224 bases at least Q40
Consensus quality: 118449 bases at least Q30
Consensus quality: 118634 bases at least Q20
Insert size: 119629; sum-of-contigs
Insert size: 117431; 9.3% error; agarose-fp
Quality coverage: 18.41x in Q20 bases; sum-of-contigs Quality
coverage: 19.36x in Q20 bases; agarose-fp
------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1    23429: contig of 23429 bp in length
*     23430    23529: gap of 100 bp in length
*     23530    63600: contig of 40071 bp in length
*     63601    63700: gap of 100 bp in length
*     63701   100099: contig of 36399 bp in length
*    100100   100199: gap of 100 bp in length
*    100200   102366: contig of 2167 bp in length
*    102367   102466: gap of 100 bp in length
*    102467   120029: contig of 17563 bp in length.
FEATURES             Location/Qualifiers
     source          1..120029
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="p22.1-22.3"
                     /clone="RP1-282H10"
                     /clone_lib="RPCI-1"
     misc_feature    1..23429
                     /note="assembly_fragment:05837
                     clone_end:SP6
                     vector_side:left"
     misc_feature    23530..63600
                     /note="assembly_fragment:00423
                     fragment_chain:1"
     misc_feature    63701..100099
                     /note="assembly_fragment:02685
                     fragment_chain:1"
     misc_feature    100200..102366
                     /note="assembly_fragment:01789"
     misc_feature    102467..120029
                     /note="assembly_fragment:02341
                     clone_end:T7
                     vector_side:right"
ORIGIN
Query Match      99.3%;  Score 244.4;  DB 2;  Length 120029;
Best Local Similarity  99.6%;  Pred. No. 6.4e-61;
Matches  245;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy       1  GTCATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21315  GTCACTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC  21374

Qy      61  AGATTTTAAGTTTACTCCCTACTGCTGCGACCCAAGTGAAATTCCTTCTCCAGTCACCAGTGTC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21375  AGATTTTAAGTTTACTCCTACTGCTGCGACCCAAGTGAAATTCCTTCTCCAGTCACCAGTGTC  21434

Qy     121  AACCTCTACCCCCAACTGCAACGAGAGTTTTGAGGGGCATCAATCACCGAGAAGTCA  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
misc_feature     32. .494
                 /note="match: GSS: Em:B63526"
misc_feature     70. .589
                 /note="match: GSS: Em:AQ315706"
repeat_region    179. .362
                 /note="match: GSS: Em:AQ285677"
repeat_region    521. .582
                 /note="MIR repeat: matches 65. .262 of consensus"
repeat_region    720. .1022
                 /note="L2 repeat: matches 2683. .2741 of consensus"
repeat_region    1327. .1448
                 /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region    2264. .2575
                 /note="AluJo/FRAM repeat: matches 181. .298 of consensus"
repeat_region    3014. .3296
                 /note="AluSq repeat: matches 1. .311 of consensus"
misc_feature     3014. .3093
                 /note="match: STS: Em:G27290"
repeat_region    5024. .5477
                 /note="40 copies 2 mer ga 75% conserved"
repeat_region    5581. .5859
                 /note="L1M1 repeat: matches 1012. .1598 of consensus"
repeat_region    5986. .6359
                 /note="AluSg repeat: matches 18. .294 of consensus"
repeat_region    6367. .6515
                 /note="L1MEc repeat: matches 2228. .2259 of consensus"
repeat_region    6971. .7198
                 /note="FLAM_C repeat: matches 1. .143 of consensus"
repeat_region    8403. .8838
                 /note="L2 repeat: matches 1512. .1735 of consensus"
misc_feature     8799. .9441
                 /note="LTR7 repeat: matches 1. .445 of consensus"
misc_feature     complement(9393. .9853)
                 /note="match: GSS: Em:B54691"
repeat_region    10311. .10437
                 /note="match: GSS: Em:AQ285817"
repeat_region    10710. .11022
                 /note="L2 repeat: matches 2612. .2747 of consensus"
repeat_region    12006. .12156
                 /note="AluJo repeat: matches 1. .309 of consensus"
repeat_region    12182. .13011
                 /note="L2 repeat: matches 2594. .2748 of consensus"
repeat_region    13404. .13631
                 /note="L2 repeat: matches 1374. .2344 of consensus"
repeat_region    13479. .13630
                 /note="114 copies 2 mer at 55% conserved"
repeat_region    13785. .14820
                 /note="38 copies 4 mer tata 60% conserved"
repeat_region    14928. .16008
                 /note="L1ME repeat: matches 4766. .5798 of consensus"
repeat_region    16029. .16110
                 /note="L1PB3 repeat: matches 5106. .6149 of consensus"
repeat_region    16428. .16829
                 /note="L1PB3 repeat: matches 5054. .5133 of consensus"
repeat_region    16868. .17150
                 /note="L1MC3 repeat: matches 6189. .6603 of consensus"
repeat_region    17246. .17991
                 /note="AluSx repeat: matches 1. .285 of consensus"
repeat_region    17992. .18675
                 /note="L1MC4 repeat: matches 6608. .7343 of consensus"
repeat_region    18676. .18729
                 /note="MER67C repeat: matches 1. .710 of consensus"
repeat_region    18731. .19458
                 /note="L1MC4 repeat: matches 7343. .7396 of consensus"
misc_feature     complement(18929. .19542)
                 /note="MER4D repeat: matches 232. .973 of consensus"
misc_feature     complement(19031. .19542)
                 /note="match: GSS: Em:AQ342961"
misc_feature     complement(19065. .19495)
                 /note="match: GSS: Em:AQ053552"
misc_feature     complement(19130. .19546)
                 /note="match: GSS: Em:AQ545892"
misc_feature     /note="match: GSS: Em:B94529"

misc_feature     19265. .19495
                 /note="match: GSS: Em:AQ373351 Em:AQ375803
                 match: STS: Em:G59266"
misc_feature     complement(19269. .19546)
                 /note="match: GSS: Em:AQ663911"
misc_feature     complement(19325. .19546)
                 /note="match: GSS: Em:AQ021494"
misc_feature     complement(19337. .19501)
                 /note="match: GSS: Em:AQ268095"
misc_feature     complement(19370. .19546)
                 /note="match: GSS: Em:AQ128461"
repeat_region    19566. .19739
                 /note="AluSg/x repeat: matches 126. .295 of consensus"
repeat_region    19740. .20100
                 /note="L1MD3 repeat: matches 7391. .7739 of consensus"
repeat_region    21121. .21429
                 /note="AluJo repeat: matches 1. .310 of consensus"
repeat_region    22589. .22728
                 /note="70 copies 2 mer aa 60% conserved"
repeat_region    22591. .22670
                 /note="20 copies 4 mer aaag 80% conserved"
repeat_region    22686. .22761
                 /note="19 copies 4 mer aagg 93% conserved"
repeat_region    23098. .23393
                 /note="AluJb repeat: matches 4. .298 of consensus"
repeat_region    23468. .23578
                 /note="L2 repeat: matches 2580. .2700 of consensus"
repeat_region    23980. .24292
                 /note="AluSq repeat: matches 2. .310 of consensus"
misc_feature     complement(27208. .27677)
                 /note="match: GSS: Em:AQ702871"
repeat_region    27819. .27957
                 /note="AluJb repeat: matches 163. .299 of consensus"
repeat_region    28701. .28930
                 /note="MLT1G repeat: matches 32. .301 of consensus"
repeat_region    29080. .29171
                 /note="MLT1H repeat: matches 433. .526 of consensus"
repeat_region    29303. .29711
                 /note="L2 repeat: matches 2323. .2710 of consensus"
repeat_region    29712. .29834
                 /note="MLT1G repeat: matches 29. .147 of consensus"
repeat_region    30140. .30547
                 /note="L2 repeat: matches 1806. .2252 of consensus"
misc_feature     complement(30720. .31207)
                 /note="match: STS: Em:HS1DJ17T"
repeat_region    30919. .31120
                 /note="MER58A repeat: matches 1. .208 of consensus"
repeat_region    32109. .32472
                 /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region    33360. .33511
                 /note="MIR repeat: matches 98. .250 of consensus"
repeat_region    34031. .34331
                 /note="AluSx repeat: matches 1. .297 of consensus"
repeat_region    34621. .34732
                 /note="MIR repeat: matches 120. .232 of consensus"
repeat_region    34888. .35007
                 /note="L2 repeat: matches 1836. .1959 of consensus"
repeat_region    35164. .35461
                 /note="AluSq repeat: matches 2. .298 of consensus"
repeat_region    35532. .35990
                 /note="L2 repeat: matches 2254. .2710 of consensus"
repeat_region    36631. .36790
                 /note="L2 repeat: matches 2342. .2501 of consensus"
repeat_region    37251. .37544
                 /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region    38446. .38493
                 /note="12 copies 4 mer caca 75% conserved"

Query Match         99.3%;  Score 244.4;  DB 9;  Length 152966;
Best Local Similarity 99.6%;  Pred. No. 6.5e-61;
Matches 245;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy   1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTAAGATTCCTGTGTTTATACAC 60
```

```
Db     53689  GTCACTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTTGTTGTTTATACAC  53748
                     |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy        61  AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC   120
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     53749  AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC  53808
Qy       121  AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA    180
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     53809  AACCTCTACCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA   53868
Qy       181  CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC   240
                     |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db     53869  CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC  53928
Qy       241  ACACTA   246
                     ||||||
Db     53929  ACACTA  53934
```

```
RESULT 7
AC106070/c
LOCUS       AC106070            211530 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-128D5, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC106070
VERSION     AC106070.5  GI:30521767
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 211530)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.I., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,W., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,M.,
            Miloavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willison,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,M., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 211530)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 211530)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24819565.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            --------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ----------------- Project Information
              Center project name: GHGQ
              Center clone name: CH230-128D5
            ------------------ Summary Statistics
              Assembly program: Atlas 3.0;
              Consensus quality: 195751 bases at least Q40
              Consensus quality: 198846 bases at least Q30
              Consensus quality: 200700 bases at least Q20
              Estimated insert size: 204266; sum-of-contigs estimation
              Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 2 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            *   This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *       1    210290: contig of 210290 bp in length
            *   210291    210390: gap of unknown length
            *   210391    211530: contig of 1140 bp in length.
FEATURES             Location/Qualifiers
     source          1..211530
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-128D5"
     misc_feature    1..1797
```

```
                    /note="wgs end extension
                    clone_end:Sp6"
misc_feature        1848..3667
                    /note="wgs end extension
                    clone_end:Sp6"
misc_feature        4029..4885
                    /note="clone boundary
                    clone_end:Sp6
                    site:EcoRI
                    end_sequence:BH279609"
misc_feature        175368..176614
                    /note="wgs_contig"
misc_feature        complement(209004..209854)
                    /note="clone boundary
                    clone_end:T7
                    site:EcoRI
                    end_sequence:BH279605"
ORIGIN

Query Match          16.2%;     Score 39.8;  DB 2;  Length 211530;
Best Local Similarity 50.3%;    Pred. No. 1.1;
Matches 98;  Conservative  0;  Mismatches 97;  Indels  0;  Gaps  0;

Qy      36 TTTAAGATTCCTGTGTTGTTATACACAGATTTAAGTTTACTCCTACTGCTGACCCAAGTG 95
              |||||||  |||||    ||||  ||||||||    ||||||||  ||  || || |||
Db  116902 TCTTAGCTTCATCAATCTTATCTACGTTTGGTGGCTGTATTTATATGGGCCACATGTGAG 116843

Qy      96 AAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCCCAACTGCAACGAGAGTTTGAG 155
              |||||    |||||  |  |||    || | ||  |  ||||||  |  |||| |||||
Db  116842 GCAGGCTCTGAAATAGTCGTTCCTTCAGTCTCTGCTCCGAACTTGAGACAAGTAATCTTAA 116783

Qy     156 GGGCATCAATCACACCGAGAAGTCACAGCCCCTCAACCACTGAGGTGTGGGGGGGTAGGG 215
              ||||  | ||||  |||    ||    ||||||  ||  ||||||||  ||||||||
Db  116782 TGACCTTAAGCTCACTGAGAAGGCTAGGCTAACCACTGAAGCTTAGGCATCCCGCT 116723

Qy     216 ATCTGCATTTCTTCA 230
              ||||| ||| ||||||
Db  116722 ATCCGCACTTCTCCA 116708

RESULT 8
AC122565/c
LOCUS       AC122565        66522 bp    DNA    linear    HTG 24-NOV-2002
DEFINITION  Mus musculus clone RP24-239A21, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC122565
VERSION     AC122565.2  GI:25229256
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 66522)
AUTHORS     Birren,B., Nusbaum,C. and Lander,E.
TITLE       Direct Submission
JOURNAL     Mus musculus, clone RP24-239A21
            Unpublished
REFERENCE   2  (bases 1 to 66522)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL     Direct Submission
            Submitted (24-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 66522)
AUTHORS     Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Major,J.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (24-NOV-2002) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 24, 2002 this sequence version replaced gi:21166330.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L26591
            Center clone name: 239_A_21
            -----------------
          * NOTE: This record contains 83 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *       1:      contig of 652 bp in length
          *     653:      gap of 100 bp
          *     753:      contig of 707 bp in length
          *    1460:      gap of 100 bp
          *    1559:      contig of 724 bp in length
          *    2283:      gap of 100 bp
          *    2384:      contig of 696 bp in length
          *    3079:      gap of 100 bp
          *    3180:      contig of 699 bp in length
          *    3879:      gap of 100 bp
          *    3978:      contig of 706 bp in length
          *    4684:      gap of 100 bp
          *    4784:      contig of 706 bp in length
          *    5490:      gap of 100 bp
          *    5591:      contig of 681 bp in length
          *    6272:      gap of 100 bp
          *    6372:      contig of 703 bp in length
          *    7075:      gap of 100 bp
          *    7175:      contig of 712 bp in length
```

```
* 7887    7986:  gap of 100 bp
* 7987    8705:  contig of 719 bp in length
* 8706    8805:  gap of 100 bp
* 8806    9511:  contig of 706 bp in length
* 9512    9611:  gap of 100 bp
* 9612   10321:  contig of 710 bp in length
* 10322  10421:  gap of 100 bp
* 10422  11116:  contig of 695 bp in length
* 11117  11216:  gap of 100 bp
* 11217  11917:  contig of 701 bp in length
* 11918  12017:  gap of 100 bp
* 12018  12726:  contig of 709 bp in length
* 12727  12826:  gap of 100 bp
* 12827  13529:  contig of 703 bp in length
* 13530  13629:  gap of 100 bp
* 13630  14336:  contig of 707 bp in length
* 14337  14436:  gap of 100 bp
* 14437  15153:  contig of 717 bp in length
* 15154  15253:  gap of 100 bp
* 15254  15953:  contig of 700 bp in length
* 15954  16053:  gap of 100 bp
* 16054  16747:  contig of 694 bp in length
* 16748  16847:  gap of 100 bp
* 16848  17572:  contig of 725 bp in length
* 17573  17672:  gap of 100 bp
* 17673  18382:  contig of 710 bp in length
* 18383  18482:  gap of 100 bp
* 18483  19190:  contig of 708 bp in length
* 19191  19290:  gap of 100 bp
* 19291  19979:  contig of 699 bp in length
* 19980  20079:  gap of 100 bp
* 20080  20777:  contig of 689 bp in length
* 20777  20877:  gap of 100 bp
* 20877  21581:  contig of 698 bp in length
* 21581  21681:  gap of 100 bp
* 21681  22377:  contig of 704 bp in length
* 22377  22477:  gap of 100 bp
* 22477  23183:  contig of 696 bp in length
* 23183  23283:  gap of 100 bp
* 23283  23986:  contig of 706 bp in length
* 23986  24086:  gap of 100 bp
* 24086  24793:  contig of 703 bp in length
* 24793  24893:  gap of 100 bp
* 24893  25592:  contig of 707 bp in length
* 25592  25692:  gap of 100 bp
* 25692  26417:  contig of 699 bp in length
* 26417  26517:  gap of 100 bp
* 26517  27231:  contig of 725 bp in length
* 27231  27331:  gap of 100 bp
* 27331  28048:  contig of 714 bp in length
* 28048  28148:  gap of 100 bp
* 28148  28843:  contig of 717 bp in length
* 28843  28943:  gap of 100 bp
* 28943  29638:  contig of 695 bp in length
* 29638  29738:  gap of 100 bp
* 29738  30456:  contig of 695 bp in length
* 30456  30556:  gap of 100 bp
* 30556  31234:  contig of 718 bp in length
* 31234  31334:  gap of 100 bp
* 31334  32037:  contig of 678 bp in length
* 32037  32137:  gap of 100 bp
* 32137  32846:  contig of 703 bp in length
* 32846  32946:  gap of 100 bp
* 32946  33664:  contig of 709 bp in length
* 33664  33764:  gap of 100 bp
* 33764  34478:  contig of 718 bp in length
* 34478  34578:  gap of 100 bp
* 34578  35275:  contig of 714 bp in length
* 35275  35375:  gap of 100 bp
* 35375  36084:  contig of 697 bp in length
* 36084  36184:  gap of 100 bp
* 36184  36896:  contig of 709 bp in length
* 36896  36996:  gap of 100 bp
* 36996          contig of 712 bp in length
```

```
* 36997  37686:  contig of 690 bp in length
* 37686  37686:  gap of 100 bp
* 37687  38472:  contig of 686 bp in length
* 37787  38472:  gap of 100 bp
* 38473  39274:  contig of 702 bp in length
* 38573  39274:  gap of 100 bp
* 39275  40071:  contig of 697 bp in length
* 39374  40071:  gap of 100 bp
* 40072  40869:  contig of 698 bp in length
* 40172  40869:  gap of 100 bp
* 40870  41674:  contig of 705 bp in length
* 40970  41674:  gap of 100 bp
* 41675  42453:  contig of 679 bp in length
* 41775  42453:  gap of 100 bp
* 42454  43261:  contig of 708 bp in length
* 42554  43261:  gap of 100 bp
* 43262  44073:  contig of 712 bp in length
* 43362  44073:  gap of 100 bp
* 44074  44868:  contig of 695 bp in length
* 44174  44868:  gap of 100 bp
* 44869  45662:  contig of 694 bp in length
* 44969  45662:  gap of 100 bp
* 45663  46455:  contig of 693 bp in length
* 45763  46455:  gap of 100 bp
```

```
Query Match            15.7%;   Score 38.6;  DB 2;  Length 65522;
Best Local Similarity  52.1%;   Pred. No. 2.4;
  Matches   86;  Conservative   0;  Mismatches   79;  Indels   0;  Gaps   0;

Qy      33 TTCTTTAAGATTCCTGTTGTTATACACAGATTTTAAGTTTACTCCTACTGCTGACCCAA 92
              |||||| ||||| |||| ||||||||||||||| |||||||||||||||| ||||||
Db   21504 TTCTTTCACATGCAGGCTTTCAATACCACTTATTAAATGCAACTCACAATGCAGACCCTC 21445

Qy      93 GTGAAATTCCTTCTCCCAGTGTCAACCTCTACCCCCCAACTGCAACGAGAGTTT 152
              ||||| || |||||
Db   21444 TGGCATAAACTGCTTCCAGTAAACTGCTATGCTTTCAACTTTCAGCAGTGAGACCTGA 21385

Qy     153 GAGGGGCATCAATCACACCGAGAAGTCACAGCCCCTCAACCACTG 197
              ||| ||| ||| ||| |||  ||||
Db   21384 GGCAGATATCAGTCAGCCCAGGCACTCTTGCTCCTAAACAGCAG 21340
```

```
RESULT 9
AC129208
LOCUS       AC129208              176244 bp    DNA     linear   ROD 27-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-282E2 from chromosome 15, complete
            sequence.
ACCESSION   AC129208
VERSION     AC129208.4  GI:3285229
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 176244)
  AUTHORS   Shah,N. and Bielicki,L.
  TITLE     The sequence of Mus musculus BAC clone RP24-282E2
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 176244)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 176244)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 176244)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-APR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 176244)
  AUTHORS   Wilson,R.K.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (26-JUL-2003) Genome Sequencing Center, 4444 Forest Park
          Parkway, St. Louis, MO 63108, USA
REFERENCE 6  (bases 1 to 176244)
AUTHORS   Wilson,R.
TITLE     Direct Submission
JOURNAL   Submitted (27-NOV-2003) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Jul 26, 2003 this sequence version replaced gi:30089830.
          ------------- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu
          Contact: submissions@watson.wustl.edu
          ------- Summary Statistics
          Center project name: M_BB0282E02
          .

NOTICE:   This sequence may not represent the entire insert of this
clone.  It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC098719.

FEATURES             Location/Qualifiers
     source          1..176244
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP24-282E2"
                     /clone_lib="RPCI-24"
     repeat_region   3420..3647
                     /rpt_family="MIR"
     repeat_region   4355..4594
                     /rpt_family="ERVK"
     repeat_region   5094..5356
                     /rpt_family="ERVK"
     repeat_region   5399..6422
                     /rpt_family="ERVK"
     repeat_region   6416..7025
                     /rpt_family="ERVK"
     repeat_region   11493..11743
                     /rpt_family="L1"
     repeat_region   11743..12500
                     /rpt_family="L1"
     repeat_region   12521..12903
                     /rpt_family="MaLR"
     repeat_region   14087..14178
                     /rpt_family="ERV1"
     repeat_region   14603..14907
                     /rpt_family="ERVK"
     repeat_region   17163..17284
                     /rpt_family="ERVK"
     repeat_region   18154..18606
                     /rpt_family="L1"
     repeat_region   18667..18752
                     /rpt_family="L1"
     repeat_region   18845..18955
                     /rpt_family="Alu"
     repeat_region   19680..19899
                     /rpt_family="B4"
     repeat_region   20285..20638
                     /rpt_family="MaLR"
     repeat_region   21034..21106
                     /rpt_family="ID"
     repeat_region   21673..21740
                     /rpt_family="ERV1"
     repeat_region   22063..22324
                     /rpt_family="ERVL"
     repeat_region   22514..22980
                     /rpt_family="L1"
     repeat_region   23845..24018
                     /rpt_family="L2"
     repeat_region   25126..25605
                     /rpt_family="ERVK"
     repeat_region   25722..25880
                     /rpt_family="ERVK"
     repeat_region   25856..25961
                     /rpt_family="ERVK"
     repeat_region   25937..26593
                     /rpt_family="ERVK"
     repeat_region   26594..26874
                     /rpt_family="ERVK"
     repeat_region   26871..27626
                     /rpt_family="ERVK"
     repeat_region   27629..28130
                     /rpt_family="ERVK"
     repeat_region   28459..28693
                     /rpt_family="L2"
     repeat_region   29279..29640
                     /rpt_family="L2"
     repeat_region   31661..31840
                     /rpt_family="L2"
     repeat_region   31877..32023
                     /rpt_family="Alu"
     repeat_region   32303..32357
                     /rpt_family="ERV1"
     repeat_region   32399..32896
                     /rpt_family="L1"
     repeat_region   32908..33064
                     /rpt_family="L1"
     repeat_region   33242..33338
                     /rpt_family="MER1_type"
     repeat_region   36084..36137
                     /rpt_family="B2"
     repeat_region   36340..36756
                     /rpt_family="L1"
     repeat_region   36757..37043
                     /rpt_family="L1"
     repeat_region   37054..39354
                     /rpt_family="L1"
     repeat_region   39426..39887
                     /rpt_family="L1"
     repeat_region   39891..40039
                     /rpt_family="ERVK"
     repeat_region   40040..40114
                     /rpt_family="ERVK"
     repeat_region   40328..40975
                     /rpt_family="ERVK"
     repeat_region   42476..43157
                     /rpt_family="ERVK"
     repeat_region   43153..43264
                     /rpt_family="L1"
```

```
    repeat_region    /rpt_family="L1"
                     4273..43708
    repeat_region    /rpt_family="L1"
                     43719..44211
    repeat_region    /rpt_family="ERVK"
                     44367..44451
    repeat_region    /rpt_family="ERVK"
                     44461..44859
    repeat_region    /rpt_family="ERVK"
                     44862..45948
    repeat_region    /rpt_family="L1"
                     45936..46544
    repeat_region    /rpt_family="L1"
                     46545..46891
    repeat_region    /rpt_family="MaLR"
                     46892..47004
    repeat_region    /rpt_family="L1"
                     47010..47183
    misc_feature     /rpt_family="L1"
                     47812..48413
                     /note="Sequence derived from PCR product of project DNA."
    unsure           48492..48573
    repeat_region    53101..54440
                     /note="Sequence derived from one plasmid subclone."


Query Match           15.7%;  Score 38.6;  DB 10;  Length 176244;
Best Local Similarity 52.1%;  Pred. No. 2.5;
Matches   86;  Conservative   0;  Mismatches  79;  Indels   0;  Gaps   0;

Qy       33 TTCTTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTTACTCCTACTGCTGACCCAA 92
            |||||| ||| ||  |||  || ||| ||| ||| ||| ||| || ||| ||||||||
Db   175171 TTCTTTCACATGCAGGCTTTCAATACCCACTTATTAAATGCAACTCACAATGAGACCCTC 175230

Qy       93 GTGAAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCCCAAACTGCAACGAGAGTTTT 152
            | | | || || || |||| || ||| | |||| |   ||| || ||||||| ||
Db   175231 TGGCATAAACTGCTCCAGTAAACTGCTATGCTTTCAACTTTCAGCAGCAGTGAGACCTGA 175290

Qy      153 GAGGGCGCATCAATCACACCGAGAAGTCACAGCCCCTCAACCACTG 197
            | ||| || ||||  ||| || || ||| |||| ||||||||
Db   175291 GGCAGATCAGTCAGCCCAGGCACTCTTGCTCCCTAAACAGCAG 175335


RESULT 10
AL929176/c
LOCUS       AL929176      100902 bp    DNA    linear    VRT 09-OCT-2003
DEFINITION  Zebrafish DNA sequence from clone XX-187G17 in linkage group 3
            Contains two novel genes similar to hbaal (hemoglobin alpha-adult
            1), the bal gene for bal globin, a novel gene similar to bal, a
            novel gene for alpha and a novel gene for beta globin, a novel
            gene, a novel gene similar to human CGTHBA (conserved gene
            telomeric to alpha globin cluster), a novel gene similar to human
            MPG (N-methylpurine-DNA glycosylase) and two CpG islands, complete
            sequence.
ACCESSION   AL929176
VERSION     AL929176.4  GI:24395450
KEYWORDS    HTG; alpha globin; bal; beta globin; CGTHBA; CpG island; DNA
            glycosylase; hbaal; hemoglobin; MPG.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 100902)
AUTHORS     Phillimore,B.
TITLE       Direct Submission
JOURNAL     Submitted (06-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 25, 2002 this sequence version replaced gi:23395900.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep XX-187G17 is
            from a Incyte Zebrafish AB BAC library VECTOR: pBeloBAC11
            ------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            Clone-derived Zebrafish pUC subclones occasionally display
            inconsistency over the length of mononucleotide A/T runs and
            conserved TA repeats. Where this is found the longest good quality
            representation will be submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU). For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml This
            sequence is the entire insert of clone XX-187G17.
FEATURES             Location/Qualifiers
     source          1..100902
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="XX-187G17"
                     /clone_lib="Incyte-BAC"
     repeat_region   2..51
                     /note="6.7 copies 3 mer AGA 40% conserved"
     repeat_region   52..71
                     /note="Dr000309 repeat: matches 799..849 of consensus"
     repeat_region   complement(75..129)
                     /note="drr806 repeat: matches 1..56 of consensus"
     repeat_region   454..464
                     /note="11.0 copies 1 mer T 22% conserved"
     repeat_region   690..703
                     /note="2.8 copies 5 mer AAAAT 28% conserved"
     repeat_region   735..956
                     /note="drr897 repeat: matches 1..228 of consensus"
     repeat_region   971..1314
                     /note="drr1203 repeat: matches 1..336 of consensus"
     repeat_region   1404..1691
                     /note="Dr000063 repeat: matches 292..592 of consensus"
     repeat_region   1687..1981
                     /note="TDR1 repeat: matches 938..1236 of consensus"
     repeat_region   1992..2009
                     /note="9.0 copies 2 mer TA 20% conserved"
     repeat_region   complement(2013..2210)
                     /note="Dr000218 repeat: matches 3..258 of consensus"
     repeat_region   2072..2252
                     /note="Dr000400 repeat: matches 407..592 of consensus"
     repeat_region   2373..2912
                     /note="Dr000400 repeat: matches 744..1298 of consensus"
     repeat_region   2909..3015
                     /note="Dr000218 repeat: matches 355..467 of consensus"
     repeat_region   3013..3368
                     /note="Dr000396 repeat: matches 4..362 of consensus"
     repeat_region   complement(3195..3394)
                     /note="Dr000218 repeat: matches 181..384 of consensus"
     repeat_region   3374..3477
                     /note="Dr000218 repeat: matches 454..561 of consensus"
```

```
repeat_region   3482..3493
                /note="2.4 copies 5 mer TATTT 24% conserved"
repeat_region   complement(3766..3852)
                /note="TDR23 repeat: matches 340..431 of consensus"
repeat_region   4055..4067
                /note="13.0 copies 1 mer T 26% conserved"
repeat_region   4234..4256
                /note="2.1 copies 11 mer AATTCTGTGAT 39% conserved"
repeat_region   4301..4317
                /note="2.4 copies 7 mer GAAAACA 25% conserved"
repeat_region   4375..4387
                /note="13.0 copies 1 mer T 26% conserved"
repeat_region   4516..4527
                /note="2.0 copies 6 mer TTATAT 24% conserved"
repeat_region   4725..4735
                /note="2.2 copies 5 mer ATTTT 22% conserved"
repeat_region   4782..4796
                /note="2.1 copies 7 mer ACAAAT 30% conserved"
repeat_region   complement(4959..5177)
                /note="drr554 repeat: matches 10..231 of consensus"
repeat_region   5208..5223
                /note="4.0 copies 4 mer ATAA 23% conserved"
repeat_region   5276..5292
                /note="2.8 copies 6 mer ATTTTG 25% conserved"
repeat_region   5650..5669
                /note="2.2 copies 9 mer AAAATTCGC 40% conserved"
repeat_region   5697..5706
                /note="10.0 copies 1 mer T 20% conserved"
repeat_region   5941..5953
                /note="2.2 copies 6 mer AGTCTG 26% conserved"
repeat_region   6207..6225
                /note="3.2 copies 6 mer AAGTTT 31% conserved"
repeat_region   6518..6530
                /note="3.2 copies 4 mer TATT 26% conserved"
gene            complement(6603..7431)
                /gene="SI:bY187G17.1"
mRNA            complement(join(6603..6836,6934..7141,7289..7431))
                /gene="SI:bY187G17.1"
                /product="SI:bY187G17.1 (novel protein similar to
                zebrafish hemoglobin alpha-adult 1 (hbaa1) )"
                /note="match: cDNAs: Em:M25643 Em:AB015448 Em:D88113
                Em:AB080119
                match: ESTs: Em:BM957615 Em:AW175412 Em:BM861123
                Em:BM776247 Em:BM775576 Em:BQ092377 Em:BI866156
                Em:BI473146 Em:BM070914 Em:BM185045 Em:BM777875
                Em:BI840610"
                /evidence=not experimental
polyA_site      complement(6603)
                /gene="SI:bY187G17.1"
polyA_site      complement(6607)
                /gene="SI:bY187G17.1"
polyA_signal    complement(6622..6627)
                /gene="SI:bY187G17.1"
CDS             complement(join(6708..6836,6934..7141,7289..7383))
                /gene="SI:bY187G17.1"
                /note="match: proteins: Sw:Q90487 Tr:O13136 Tr:O13139"
                /codon_start=1
                /evidence=not experimental
                /product="SI:bY187G17.1 (novel protein similar to
                zebrafish hemoglobin alpha-adult 1 (hbaa1) )"
                /protein_id="CAE48980.1"
                /db_xref="GI:37606095"
                /translation="MSLSDKDKAVVKAIWAKISPKADEIGAEALARMLTVYPQTKTYF
                SHWSDLSPGSGPVKKHGKTIMGAVGEAISKIDDLVGGLAALSELHAFKLRVDPANFKI
                LSHNVIVIAMLFPADFTPEVHVSVDKFFNNLALALSEKYR"
repeat_region   6864..6874
                /note="2.2 copies 5 mer TACCA 22% conserved"
repeat_region   6910..6921
                /note="2.4 copies 5 mer TGCTG 24% conserved"
repeat_region   7210..7221
                /note="4.0 copies 3 mer TAA 24% conserved"
repeat_region   7218..7237
                /note="2.0 copies 10 mer ATAAACTTCT 31% conserved"
```

```
repeat_region   7847..7856
                /note="2.5 copies 4 mer AACA 20% conserved"
repeat_region   7940..8016
                /note="25.7 copies 3 mer TTA 118% conserved"
repeat_region   8107..8116
                /note="10.0 copies 1 mer A 20% conserved"
gene            8294..9109
                /gene="bal"
mRNA            join(8294..8442,8543..8765,8865..9109)
                /gene="bal"
                /product="SI:bY187G17.2 (bal globin)"
                /note="ZFIN record: ZDB-GENE-990415-18
                match: cDNAs: Em:U50380
                match: ESTs: Em:BM957208 Em:AW280571 Em:BQ449732
                Em:BI428616 Em:BQ419496 Em:BI842728 Em:BI840804
                Em:BI980962 Em:BM156057 Em:BI840865 Em:BI534296
                Em:BM184992 Em:BI705778"
                /evidence=not experimental
CDS             join(8351..8442,8543..8765,8865..8996)
                /gene="bal"
                /note="ZFIN record: ZDB-GENE-990415-18
                match: proteins: Tr:Q90486"
                /codon_start=1
                /evidence=not experimental
                /product="SI:bY187G17.2 (bal globin)"

Query Match             15.4%;  Score 37.8;  DB 5;  Length 100902;
Best Local Similarity   52.9%;  Pred. No. 4.2;
Matches  81;  Conservative   0;  Mismatches  72;  Indels   0;  Gaps   0;

Qy      2 TCATTAACCTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTATACACA 61
          |||||| ||| || |||||| | ||   |||| ||||   || |||||| ||| || ||||
Db  60729 TAAGAATCCATTTTATTTTATTTTATTCATTAATTTTACCTGTTTATTGGTGCATATATA 60670

Qy     62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA 121
          |||| |||||| || || ||  |  ||| |||  |||  || || || ||  | ||||
Db  60669 GGCTATAATTAACTCATACAACAGTACCTTGGCAGCAGTCACGTGCCACTAAAGCTTTCA 60610

Qy    122 ACCTCTACCCCCCAACTGCAACGAGAGTTTGA 154
          ||||| ||  ||  |||| | ||  || ||||
Db  60609 AACACAACCCCATAAAGTAATGAAAGATTAGA 60577


RESULT 11
AC121787/c
LOCUS       AC121787              200985 bp    DNA     linear   ROD 11-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-393K24 from chromosome 9, complete
            sequence.
ACCESSION   AC121787
VERSION     AC121787.3  GI:27413943
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200985)
  AUTHORS   Harris,A., Holmes,A. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP23-393K24
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 200985)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 200985)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 200985)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
```

```
REFERENCE   5  (bases 1 to 200985)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (29-DEC-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 200985)
AUTHORS     Wilson,R.
TITLE       Direct Submission
JOURNAL     Submitted (11-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Dec 29, 2002 this sequence version replaced gi:22475661.
            ---------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            --------- Summary Statistics
            Center project name: M_BA0393K24
            --------------

            NOTICE: This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA. The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..200985
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="9"
                     /clone="RP23-393K24"
                     /clone_lib="RPCI-23"
     repeat_region   41..171
                     /rpt_family="L1"
     repeat_region   841..1023
                     /rpt_family="B2"
     repeat_region   2407..2565
                     /rpt_family="B4"
     repeat_region   2549..2964
                     /rpt_family="ERVK"
     repeat_region   3120..3418
                     /rpt_family="MaLR"
     repeat_region   3531..3744
                     /rpt_family="MaLR"
     repeat_region   5631..5771
                     /rpt_family="B4"
     repeat_region   5794..5920
                     /rpt_family="B4"
     repeat_region   6038..6360
                     /rpt_family="MaLR"
     repeat_region   6637..6777
                     /rpt_family="B4"
     repeat_region   6834..7034
                     /rpt_family="MaLR"
     repeat_region   7025..7076
                     /rpt_family="MER1_type"
     repeat_region   7077..7279
                     /rpt_family="B2"
     repeat_region   7280..7410
                     /rpt_family="MER1_type"
     repeat_region   7758..8473
                     /rpt_family="L1"
     repeat_region   8549..8658
                     /rpt_family="Alu"
     repeat_region   8749..8884
                     /rpt_family="B4"
     repeat_region   9507..9644
                     /rpt_family="B4"
     repeat_region   9734..10063
                     /rpt_family="L1"
     repeat_region   10070..10165
                     /rpt_family="L1"
     repeat_region   10166..10280
                     /rpt_family="Alu"
     repeat_region   10319..10697
                     /rpt_family="L1"
     repeat_region   10698..11080
                     /rpt_family="MaLR"
     repeat_region   11081..11678
                     /rpt_family="L1"
     repeat_region   11822..11926
                     /rpt_family="L1"
     repeat_region   13187..13310
                     /rpt_family="Alu"
     repeat_region   13624..13947
                     /rpt_family="MaLR"
     repeat_region   14694..15518
                     /rpt_family="L1"
     repeat_region   15784..15869
                     /rpt_family="Alu"
     repeat_region   18063..19022
                     /rpt_family="L1"
     repeat_region   19186..19462
                     /rpt_family="L1"
     repeat_region   19466..19714
                     /rpt_family="L1"
     repeat_region   19705..19827
                     /rpt_family="B4"
     repeat_region   19846..20018
                     /rpt_family="L1"
     repeat_region   21554..22091
                     /rpt_family="MaLR"
     repeat_region   20045..20625
                     /rpt_family="MaLR"
     repeat_region   22161..22558
                     /rpt_family="ERVK"
     repeat_region   21099..22236
                     /rpt_family="L1"
     repeat_region   22631..23093
                     /rpt_family="MaLR"
     repeat_region   21282..21394
                     /rpt_family="L1"
     repeat_region   23094..23173
                     /rpt_family="MaLR"
     repeat_region   21448..21548
                     /rpt_family="B4"
     tRNA            join(23174..23210,23224..23259)
                     /product="tRNA-Arg"
     repeat_region   23174..23259
                     /rpt_family="tRNA-Arg-AGA"
```

```
repeat_region    23268..23415
                 /rpt_family="Alu"
repeat_region    23416..23760
                 /rpt_family="MaLR"
repeat_region    23764..24991
                 /rpt_family="L1"
repeat_region    25082..25241
                 /rpt_family="L1"
repeat_region    25235..25804
                 /rpt_family="L1"
repeat_region    25787..25853
                 /rpt_family="L1"
repeat_region    26147..26377
                 /rpt_family="ERVK"
repeat_region    26380..26746
                 /rpt_family="ERVK"
repeat_region    26768..26912
                 /rpt_family="Alu"
repeat_region    27005..27331
                 /rpt_family="L1"
repeat_region    27584..27776
                 /rpt_family="B2"
repeat_region    27805..27875
                 /rpt_family="ID"
repeat_region    27891..28013
                 /rpt_family="B4"
```

```
Query Match            15.3%;     Score 37.6;    DB 10;    Length 200985;
Best Local Similarity  50.0%;     Pred. No. 5;
Matches   94;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps     0;

Qy       25 TTTTTATTTCTTTAAGATTCCTGTGTTTATACACAGATTTAAGTTACTCCTACTGC 84
            |||||| ||||||| |||||||||||||||||||||||||||||||||||||||||
Db    23825 TTTTTTTCTCTTGAGGATTTCAAATAGCTTCTTCTTTGTTTTTGTATATGTAGTGTTTTAAC 23766

Qy       85 TGACCCAAGTGAAAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCCCAACTGCAACG 144
            ||||  ||||||||| ||||| ||   |||  |||| || |||||||||| ||| ||||||
Db    23765 TAACATCTTGTTGAGCTTTCATTGTTATGATAAAAATACCATGACCACAGTTGCTAAG 23706

Qy      145 AGAGTTTGACGGGCATCAATCACACCGAGAAGTCACAGCCCCTCAACCACTGAGGTGTG 204
            |||||||||||| ||| ||| ||  ||||   |||||||||| ||||||||||||| |||
Db    23705 AGAAAAAGGGGGTTTTCACTTCCACTTCTGTATCACAGTCCATCACTGAAGGAAGTCAG 23646

Qy      205 GGGGGGTA 212
            |||| |||
Db    23645 GGCAGGAA 23638
```

```
RESULT 12
LOCUS       AC129445
DEFINITION  Rattus norvegicus clone CH230-13P17, WORKING DRAFT SEQUENCE.
ACCESSION   AC129445
VERSION     AC129445.6   GI:30521811
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 242754)
  AUTHORS   Muzny,D.Marie., Metzker,M.lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,A., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,J., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,R., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic.A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,J., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 242754)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 242754)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24818813.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ------- Project Information
              Center project name: GEZX
              Center clone name: CH230-13P17
            ------- Summary Statistics
```

```
              Assembly program: Atlas 3.0;
              Consensus quality: 228124 bases at least Q40
              Consensus quality: 230998 bases at least Q30
              Consensus quality: 233152 bases at least Q20
              Estimated insert size: 237219; sum-of-contigs estimation
              Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
              --------------
           *  NOTE: Estimated insert size may differ from sequence length
           *  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
           *  NOTE: This is a 'working draft' sequence. It currently
           *  consists of 1 contigs. Gaps between the contigs
           *  are represented as runs of N. The order of the pieces
           *  is believed to be correct as given, however the sizes
           *  of the gaps between them are based on estimates that have
           *  provided by the submittor.
           *  This sequence will be replaced
           *  by the finished sequence as soon as it is available and
           *  the accession number will be preserved.
           *         1  242754: contig of 242754 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..242754
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-13P17"
     misc_feature    524..1247
                     /note="clone boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH276372"
     misc_feature    98074..99377
                     /note="wgs_contig"
     misc_feature    complement(234009..234650)
                     /note="clone boundary
                     complement(234009..234650)
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    241197..242754
                     /note="wgs_end_extension
                     clone_end:T7"
```

```
ORIGIN

Query Match           15.3%;      Score 37.6;    DB 2;   Length 242754;
Best Local Similarity 48.6%;    Pred. No. 5;
Matches   109;      Conservative    0;   Mismatches  109;   Indels  0;   Gaps    0;

Qy      9 CCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCGTCGTTGTTATACACAGAGTTTA  68
          |||||| |||| || |||||||   || || |   ||||||||| |||   ||   ||||
Db 211568 CCTGTGCTAAGAAGTGTTTTGATTTTGTTTTTTATTTCAGTCAACTAAAAAGCTTCTGC 211627

Qy     69 AGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTTCCAGTCACAGTGTCAACCTCTA  128
          ||  | || ||  |  |||  || |||||||  | |||||||||| |||||||  ||||||
Db 211628 CATTCCTGCCAACTTCTAAACCTAAGTGGTCTCCTTCCTCTCAAATAAATGGTTCTAAAGCTG 211687

Qy    129 CCCCCAACTGCAACGAGCAGAGTTTTGAGGGGCCATCAATCACACCGAGAAGTCACAGCCCCT  188
          |||||  |||  || || ||| |||   |||  ||  |||| ||||||||||  ||| ||||
Db 211688 TTTCTTGAGTCCAGATAGACTTTATTATGAGAACCATTAGAGATTTTAAAGCCCCCTCGCCT 211747

Qy    189 CAACCACTGAGGTGTGGGGGGGTAGGGGGATCTG  220
          ||||||| ||| ||| ||||||||| ||| |||
Db 211748 GTACAAGGTGGGCCAGTGTAACAGTTAGCTG 211779
```

```
RESULT 13
AC105857/c
LOCUS       AC105857      247480 bp    DNA     linear   HTG 22-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-95G8, *** SEQUENCING IN PROGRESS ***,
            2 unordered pieces.
ACCESSION   AC105857
VERSION     AC105857.4  GI:23101480
```

```
Qy      129  CCCCCAACTGCAACAGAGAGTTTTGAGGGCATCAATCACCGAGAGTCAGAGCCCCT  188
             ||||| ||||||||||||||| | ||||| |||||| ||||||||||| |||||| |||
Db   196223  TTTCTTGAGTCCAGATAGACTTATTATGAGAACCATTAGATTTAAAGCCCTTCGCCT  196164

Qy      189  CAACCACTGAGGTGTCGGGGGTAGGGATCTG  220
             |||||||||||||||||  |||| |||||
Db   196163  GTACAGGTGGGGCGAGTGTAACAGTTAGCTG  196132
```

RESULT 14
AC090127

```
LOCUS         AC090127              212404 bp    DNA    linear    ROD 27-APR-2002
DEFINITION    Mus musculus chromosome 6, clone RP23-128D23, complete sequence.
ACCESSION     AC090127
VERSION       AC090127.11  GI:20030898
KEYWORDS      HTG.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 212404)
  AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE       Mus musculus chromosome 6, clone RP23-128D23
  JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 212404)
  AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
              Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
              Camarata,J., Campopiano,A., Choepel,A., Colangelo,M., Collins,S.,
              Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
              Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
              Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
              Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
              Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
              Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
              Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
              McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
              Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
              O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
              Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
              Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
              Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
              Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
              Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
              Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
              Zembek,L., Zimmer,A. and Zody,M.
  TITLE       Direct Submission
  JOURNAL     Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 212404)
  AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
              Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
              Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
              Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
              Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
              Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
              Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
              Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
              Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
              Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
              MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
              McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
              Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
              Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
              Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
              Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Schupback,R.,
              Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
              Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
              Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
              Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
              Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
              Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE       Direct Submission
```

(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
--------------- Genome Center
      Center: Baylor College of Medicine
      Center code: BCM
      Web site: http://www.hgsc.bcm.tmc.edu/
      Contact: hgsc-help@bcm.tmc.edu
--------------- Project Information
      Center project name: GNXK
      Center clone name: CH230-95G8
--------------- Summary Statistics
      Assembly program: Phrap; version 0.990329
      Consensus quality: 223628 bases at least Q40
      Consensus quality: 227628 bases at least Q30
      Consensus quality: 227305 bases at least Q30
      Consensus quality: 229716 bases at least Q20
      Estimated insert size: 240475; sum-of-contigs estimation
      Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1 243951: contig of 243951 bp in length
* 243952 244051: gap of unknown length
* 244052 247480: contig of 3429 bp in length.
```
FEATURES             Location/Qualifiers
     source          1..247480
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-95G8"
     misc_feature    239..1070
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH359863"
     misc_feature    complement(237189..238065)
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH359858"
     misc_feature    239489..240989
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    241930..243951
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    244052..245191
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN
```

Query Match   15.3%; Score 37.6; DB 2; Length 247480;
Best Local Similarity 48.6%; Pred. No. 5;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

```
Qy        9  CCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTA  68
             ||||| ||  |  |||| |||   | |  | | ||||| ||| |||| |||||||||
Db   196343  CCTGTGCTAAGAAGTGTTTTGATTTTGTTTTTTTATTTCAGTCAACTAAAAAGCTTCTGC  196284

Qy       69  AGTTTACTCCTACTGTGACCAAGTGAAATTCCTTCTCCAGTGTCACCTCTCA  128
             ||| ||| ||  | |  |  | |||||| |||| |  ||||| |||   ||
Db   196283  CATTCCTGCCAACTTCTAAACCTAAGTGGTCTCCTTCTCAAATAAATGGTTCTCAAAGCTG  196224
```

JOURNAL   Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 212404)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,R., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Apr 26, 2002 this sequence version replaced gi:20219124.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www.seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------------- Project Information
            Center project name: L11594
            Center clone name: 128_D_23

FEATURES   Location/Qualifiers
     source   1..212404
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="6"
                     /map="6"
                     /clone="RP23-128D23"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     repeat_region   7..50
                     /rpt_family="AT_rich"
     repeat_region   complement(528..841)
                     /rpt_family="Lx8"
     repeat_region   855..1411
                     /rpt_family="L1"
     repeat_region   1491..1527
                     /rpt_family="(CA)n"
     repeat_region   complement(3065..3365)
                     /rpt_family="Lx8"
     repeat_region   complement(4852..5300)
                     /rpt_family="Lx8"
     repeat_region   5395..5436
                     /rpt_family="(TTTA)n"
     repeat_region   5447..5491
                     /rpt_family="CT-rich"
     repeat_region   complement(5509..5795)
                     /rpt_family="Lx8"
     repeat_region   complement(5799..5851)
                     /rpt_family="B1_MM"
     repeat_region   7807..7982
                     /rpt_family="(TTTC)n"
     repeat_region   complement(7995..8272)

     repeat_region   8281..8302
                     /rpt_family="MTD"
     repeat_region   8312..8353
                     /rpt_family="AT_rich"
     repeat_region   8355..8536
                     /rpt_family=" (GAAA)n"
     repeat_region   complement(8537..10801)
                     /rpt_family="GA-rich"
     unsure   complement(9385..9389)
                     /note="<30 qual SNGL region"
     unsure   complement(9492..9497)
                     /note="<30 qual SNGL region"
     repeat_region   11578..12353
                     /rpt_family="L1"
     repeat_region   12354..12466
                     /rpt_family="(CA)n"
     repeat_region   12467..13522
                     /rpt_family="L1"
     unsure   13160..13165
                     /note="<30 qual SNGL region"
     unsure   13204..13208
                     /note="<30 qual SNGL region"
     repeat_region   14255..14463
                     /rpt_family="L1"
     repeat_region   complement(15279..15586)
                     /rpt_family="Lx8"
     repeat_region   15747..16085
                     /rpt_family="ORR1A3"
     repeat_region   16086..16201
                     /rpt_family="ORR1A-int"
     repeat_region   complement(16258..16997)
                     /rpt_family="Lx"
     repeat_region   17042..17064
                     /rpt_family="(CAAAA)n"
     repeat_region   17406..17440
                     /rpt_family="AT_rich"
     repeat_region   complement(17453..18422)
                     /rpt_family="Lx2"
     repeat_region   18423..18816
                     /rpt_family="L1"
     repeat_region   21111..21130
                     /rpt_family="(CA)n"
     repeat_region   21131..21157
                     /rpt_family="(CA)n"
     repeat_region   complement(21862..22008)
                     /rpt_family="B1_MM"
     repeat_region   22363..22399
                     /rpt_family="Lx8"
     repeat_region   22719..23842
                     /rpt_family="AT_rich"
     repeat_region   23878..24271
                     /rpt_family="L1_MM"
     repeat_region   24703..24799
                     /rpt_family="MTC"
     repeat_region   complement(25422..25540)
                     /rpt_family="(TCTA)n"
     repeat_region   complement(25692..25756)
                     /rpt_family="Lx8"
     repeat_region   25762..25782
                     /rpt_family="B1F"
     repeat_region   complement(25783..27197)
                     /rpt_family="AT_rich"
     repeat_region   27265..27368
                     /rpt_family="Lx"
                     /rpt_family="Lx5"

Query Match   15.2%; Score 37.4; DB 10; Length 212404;
Best Local Similarity   60.2%;   Pred. No. 5.7;
Matches   62; Conservative   0; Mismatches 41; Indels   0; Gaps   0;

Qy   5 TRAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTGTTATTATACACAGAT 64
             | ||| |||| | |||| || || || ||||| | ||||| ||| || || || ||||| ||
Db   17395 TTTAAGTTGTCTATTATCTTTTTAAATGTTATTAAATTATTTGATTATATACTTAGTT   17454

```
Qy      65  TTTAAGTTACTCCTACTGCTGACCCAAGTGAAAATTCCTTCTC  107
                ||| ||||||||||||||  ||||||| ||||||| ||||||
Db    17455  TTTATTACAATCAAACTGTTAACCCACTCCAAGTCCATTCTC  17497
```

```
RESULT 15
AC138731/c
LOCUS       AC138731            160699 bp    DNA     linear   HTG 28-JAN-2003
DEFINITION  Pongo pygmaeus clone CH253-404N12, WORKING DRAFT SEQUENCE, 8
            ordered pieces.
ACCESSION   AC138731
VERSION     AC138731.2  GI:27923655
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pongo pygmaeus (orangutan)
ORGANISM    Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 160699)
AUTHORS     Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Maskeri,B., Masiello,C., Maskeri,B., McDowell,J.,
            Paquirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
TITLE       NISC Comparative Sequencing Initiative
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 160699)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (16-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 160699)
AUTHORS     Green,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (28-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
            On Jan 28, 2003 this sequence version replaced gi:27764645.
COMMENT     -------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
            Center project name: ela
            Center clone name: 404N12
            --------------
            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 159180 bases at least Q40
            Consensus quality: 159652 bases at least Q30
            Consensus quality: 159822 bases at least Q20
            Insert size: 147000; agarose-fp
            Insert size: 159999; sum-of-contigs
            Quality coverage: 11.17x in Q20 bases; agarose-fp
```

```
            Quality coverage: 10.26x in Q20 bases; sum-of-contigs
            --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 8 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *     1     27629: contig of 27629 bp in length
          *     27630    27729: gap of unknown length
          *     27730    59021: contig of 31292 bp in length
          *     59022    59121: gap of unknown length
          *     59122    63918: contig of 4797 bp in length
          *     63919    64018: gap of unknown length
          *     64019    91239: contig of 27221 bp in length
          *     91240    91339: gap of unknown length
          *     91340   125098: contig of 33759 bp in length
          *    125099   125198: gap of unknown length
          *    125199   132510: contig of 7312 bp in length
          *    132511   132610: gap of unknown length
          *    132611   144517: contig of 11907 bp in length
          *    144518   144617: gap of unknown length
          *    144618   160699: contig of 16082 bp in length.
FEATURES             Location/Qualifiers
     source          1..160699
                     /organism="Pongo pygmaeus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9600"
                     /clone="CH253-404N12"
                     /clone_lib="CH253"
     misc_feature    1..27629
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    27730..59021
                     /note="assembly_fragment"
     misc_feature    59122..63918
                     /note="assembly_fragment"
     misc_feature    64019..91239
                     /note="assembly_fragment"
     misc_feature    91340..125098
                     /note="assembly_fragment"
     misc_feature    125199..132510
                     /note="assembly_fragment"
     misc_feature    132611..144517
                     /note="assembly_fragment"
     misc_feature    144618..160699
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Query Match         15.1%;   Score 37.2;  DB 2;  Length 160699;
Best Local Similarity   56.6%;  Pred. No. 6.5;
Matches    69;  Conservative   0;  Mismatches   53;  Indels   0;  Gaps   0;
```

```
Qy       2  TCATTAACCTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTATACACA  61
            | || |||| ||||||||||||  ||| |||||||||||||| |||||| |||||||
Db   89101  TAACTAAGCTTCTCAAGGCTATTCTTCATCTTCACTAATATCTTTCTACTTAATCCCT  89042

Qy      62  GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAAATTCCTTCTCCCAGTCACAGTGTCA  121
            | ||||   |   |||||| ||||  ||| |||| || |||| ||||| |  ||| |||
Db   89041  GATATTCAAAATATATTTAAAGGAGGTGGGAGAGAAATGCTTCCCATGTCACATTATAA  88982

Qy     122  AC 123
            ||
Db   88981  AC 88980
```

```
Search completed: June 10, 2004, 18:22:48
```

us-09-674-593-10.rge

Job time : 1719 secs

```
                    GenCore version 5.1.6
           Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    June 10, 2004, 15:30:58 ; Search time 344 Seconds
                                      (without alignments)
                        3037.956 Million cell updates/sec

Title:          US-09-674-593-10
Perfect score:  246
Sequence:       1 gtcattaacctttgcaagga.........ttcatatcaacccacacta 246

Scoring table:  IDENTITY NUC
                Gapop 10.0 , Gapext 1.0

Searched:       373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_29Jan04:*
         1:  geneseqn1980s:*
         2:  geneseqn1990s:*
         3:  geneseqn2000s:*
         4:  geneseqn2001as:*
         5:  geneseqn2001bs:*
         6:  geneseqn2002s:*
         7:  geneseqn2003as:*
         8:  geneseqn2003bs:*
         9:  geneseqn2003cs:*
        10:  geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                         SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	246	100.0	246	3	AAZ36649	Aaz36649 EST AA863
2	244.4	99.3	1382	3	AAZ36643	Aaz36643 Human tum
c 3	38.4	15.6	242	4	AAL10336	Aal10336 Human bre
c 4	36.2	14.7	3588	7	ACA28683	Aca28683 Prokaryot
5	35.6	14.5	507	3	AAZ80742	Aaz80742 Human col
6	35.6	14.5	24387	4	AAL07156	Aal07156 Human rep
7	35.4	14.4	2803	6	ABK35341	Abk35341 Human cDN
c 8	35.4	14.4	3644	4	AAC84450	Aac84450 Nucleotid
c 9	35.4	14.4	3710	4	AAH17830	Aah17830 Human cDN
c 10	35.4	14.4	4480	6	ABK28656	Abk28656 Human cDN
c 11	34.8	14.1	2264	9	ADE53917	Ade53917 Human pro
c 12	34.4	14.0	2239	4	ABL03611	Abl03611 Drosophil
c 13	34.4	14.0	4405	4	ABL03610	Abl03610 Drosophil
c 14	34.4	14.0	12860	4	ABL03530	Abl03530 Drosophil
15	34.2	13.9	388	6	AAK55128	Aak55128 Human imm
16	34	13.8	17213	6	ABL33482	Abl33482 Human imm
17	34	13.8	110000	6	ABQ74964_6	Continuation (7 of
c 18	33.2	13.5	3759	7	ACA28109	Aca28109 Prokaryot
c 19	32.8	13.3	16560	4	ABL17922	Abl17922 Drosophil
c 20	32.8	13.3	19165	4	ABL17898	Abl17898 Drosophil
c 21	32.4	13.2	580	6	ABQ59302	Abq59302 Human col
c 22	32.2	13.1	15674	6	ABL32362	Abl32362 Human imm
c 23	32.2	13.1	15674	6	ABL34476	Abl34476 Human met
c 24	32.2	13.1	15674	6	ABL70513	Abl70513 Chemicall
25	32.2	13.1	98690	6	ABK12169	Abk12169 Human DNA
c 26	32	13.0	480	3	ACA27596	Aca27596 Prokaryot
c 27	32	13.0	2175	3	AAC79978	Aac79978 Human sec
c 28	31.8	12.9	422	8	ACH15465	Ach15465 Human adu
29	31.8	12.9	1885	3	AAA50339	Aaa50339 Human myr
30	31.8	12.9	4346	9	ADE25645	Ade25645 Human cDN
c 31	31.6	12.8	472	8	ACH39321	Ach39321 Human foe
c 32	31.6	12.8	700	4	AAH92432	Aah92432 Human inf
33	31.6	12.8	2337	8	ADA32494	Ada32494 DNA encod
34	31.6	12.8	13158	2	AAT75288	Aat75288 Nucleotid
35	31.4	12.8	1017	6	AAX99542	Aax99542 Nucleic a
c 36	31.4	12.8	1215	6	ABL90038	Abl90038 Human pol
37	31.4	12.8	8467	6	ABL32108	Abl32108 Human imm
38	31.4	12.8	8758	6	ABL33119	Abl33119 Human imm
c 39	31.2	12.7	400	7	ABX43460	Abx43460 Bovine ES
40	31.2	12.7	437	4	AAK56796	Aak56796 Human imm
c 41	31.2	12.7	458	7	ABZ18171	Abz18171 Group III
42	31.2	12.7	468	4	AAI10244	Aai10244 Probe #17
c 43	31.2	12.7	468	4	ABA51880	Aba51880 Human foe
c 44	31.2	12.7	468	4	AAI31492	Aai31492 Probe #17
c 45	31.2	12.7	468	4	ABA21701	Aba21701 Probe #16

```
                              ALIGNMENTS

RESULT 1
AAZ36649
ID   AAZ36649 standard; cDNA; 246 BP.
XX
AC   AAZ36649;
XX
DT   22-FEB-2000  (first entry)
XX
DE   EST AA863443 which overlaps the RUR-1 antisense cDNA sequence.
XX
KW   Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;
KW   renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma; leukaemia;
KW   EST; expressed sequence tag; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958546-A1.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;  99WO-US010424.
XX
PR   13-MAY-1998;  98US-0085318P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Van Den Eynde B,  Boon-Falleur T;
XX
DR   WPI; 2000-053076/04.
XX
PT   New isolated tumor rejection antigen RUR-1 nucleic acids, used for, e.g.
PT   treatment of cancers.
XX
PS   Claim 8; Page 71; 75pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST) which
CC   overlaps the antisense cDNA sequence of human tumour rejection antigen
CC   RUR-1. The RUR-1 antisense sequence is the antisense strand of a
CC   ubiquitously expressed gene. The antisense strand codes for a polypeptide
CC   which is preferentially expressed in tumour samples and tumour-derived
CC   cells lines. The polypeptide is unrelated to any TRAP protein. The RUR-1
CC   sequence was isolated from a renal cell carcinoma line LB9211-RCC. The
CC   RUR-1 nucleic acids and polypeptides can be used for diagnosis, prognosis
CC   or treatment of a disorder characterized by the expression of a RUR-1
CC   antisense cDNA molecule or an expression product, such as cancers, e.g.
CC   renal cell carcinoma, colorectal carcinoma, melanoma, sarcoma or
```

CC leukaemia. note: although the present sequence is mentioned in claim 8,
CC it is not specifically claimed. The fragments of AAZ36643-44 which DO NOT
CC contain the present sequence are claimed
XX
SQ Sequence 246 BP; 63 A; 66 C; 44 G; 73 T; 0 U; 0 Other;

Query Match   100.0%; Score 246; DB 3; Length 246;
Best Local Similarity   100.0%; Pred. No. 7.3e-65;
Matches 246;   Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120

Qy   121 AACCTCTACCCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AACCTCTACCCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 180

Qy   181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240

Qy   241 ACACTA 246
              ||||||
Db   241 ACACTA 246

RESULT 2
AAZ36643
ID   AAZ36643 standard; cDNA; 1382 BP.
XX
AC   AAZ36643;
XX
DT   22-FEB-2000 (first entry)
XX
DE   Human tumour rejection antigen RUR-1 antisense cDNA sequence.
XX
KW   Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;
KW   renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma; leukaemia;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   primer_bind   523..547
FT   /*tag= b
FT   /note= "binding site for primer VDE119 (see AAZ36647)"
FT   CDS   738..992
FT   /*tag= a
FT   primer_bind   complement(1280..1305)
FT   /*tag= c
FT   /note= "binding site for primer VDE120 (see AAZ36648)"
XX
PN   WO995546-A1.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999; 99WO-US010424.
XX
PR   13-MAY-1998; 98US-0085318P.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Van Den Eynde B, Boon-Falleur T;
XX
DR   WPI; 2000-053076/04.
DR   P-PSDB; AAY53809.
XX
PT   New isolated tumor rejection antigen RUR-1 nucleic acids, used for, e.g.

PT   treatment of cancers.
XX
PS   Claim 4; Fig 5; 75pp; English.
XX
CC   The present sequence represents the antisense cDNA sequence of human
CC   tumour rejection antigen RUR-1. The present sequence is the antisense
CC   strand of a ubiquitously expressed gene. The antisense strand codes for a
CC   polypeptide which is preferentially expressed in tumour samples and
CC   tumour-derived cell lines. The polypeptide is unrelated to any TRAP
CC   protein. The sequence was isolated from a renal cell carcinoma line
CC   LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for
CC   diagnosis, prognosis or treatment of a disorder characterized by the
CC   expression of a RUR-1 antisense cDNA molecule or an expression product,
CC   such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,
CC   melanoma, sarcoma or leukaemia
XX
SQ   Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 U; 0 Other;

Query Match   99.3%; Score 244.4; DB 3; Length 1382;
Best Local Similarity   99.6%; Pred. No. 4.2e-64;
Matches 245;   Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 GTCATTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 60
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268 GTCACTAACCTTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTTATACAC 327

Qy   61 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   328 AGATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 387

Qy   121 AACCTCTACCCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   388 AACCTCTACCCCCCAACTGCAACGAGAGTTTGAGGGGCATCAATCACACCGAGAAGTCA 447

Qy   181 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   448 CAGCCCCTCAACCACTGAGGTGTGGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 507

Qy   241 ACACTA 246
              ||||||
Db   508 ACACTA 513

RESULT 3
AAL10336/c
ID   AAL10336 standard; cDNA; 242 BP.
XX
AC   AAL10336;
XX
DT   07-DEC-2001 (first entry)
XX
DE   Human breast cancer expressed polynucleotide 2793.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015168-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US000798.
XX
PR   14-JAN-2000; 2000US-0176077P.
PR   14-MAR-2000; 2000US-0189167P.
PR   24-MAR-2000; 2000US-0192099P.
PR   29-MAR-2000; 2000US-0193480P.
PR   15-MAY-2000; 2000US-0205230P.
PR   09-JUN-2000; 2000US-0211315P.
PR   25-JUL-2000; 2000US-0220534P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX

PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 524; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 242 BP; 89 A; 56 C; 62 G; 35 T; 0 U; 0 Other;

Query Match   15.6%; Score 38.4; DB 4; Length 242;
Best Local Similarity 56.2%; Pred. No. 0.12;
Matches   72; Conservative   0; Mismatches   56; Indels   0; Gaps   0;

Qy   24 CTTTTTATTTTCTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCCTACTG 83
          || ||| |||| ||||||||| ||||||||
Db   231 CTTTCTGTTTGGTTTTTTTTTCTTTTTCTTTTTCTATATATATTTTTGTTGAATTCTATTT 172

Qy   84 CTGACCCAAGTGAAATTCCTTCTCCAGTGACAGTGTCAACCTCTACCCCCAACTGCAAC 143
          ||||||| ||||||
Db   171 TATTTTTAATTCTCTCTTCCTCTCCAGACACAATGGCACTGCTTATCTCCGAAATGGTGT 112

Qy   144 GAGAGTTT 151
          || |||
Db   111 GATCGTCT 104

RESULT 4
ACA28683/c
ID   ACA28683 standard; DNA; 3588 BP.
XX
AC   ACA28683;
XX
DT   19-JUN-2003 (first entry)
DE   Prokaryotic essential gene #10340.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Clostridium botulinum.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI   Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU24813.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 16553; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3588 BP; 1722 A; 290 C; 624 G; 952 T; 0 U; 0 Other;

Query Match   14.7%; Score 36.2; DB 7; Length 3588;
Best Local Similarity 56.2%; Pred. No. 1.4;
Matches   68; Conservative   0; Mismatches   53; Indels   0; Gaps   0;

Qy   2 TCATTAACCTTGCAAGGATACCTTTTTATTTTCTTTAAGATTCCTGTTGTTGTTATACACA 61
          |||||| ||||| ||||||||||||||||||||||
Db   3095 TCAATAACATTCAAAAGTTCTTCTTTGACTTTACTAAATCTTCCTTTTGATTATTCATA 3036

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTGACCAGTGTCA 121
          ||||||||| |||||| |||
Db   3035 AATGTTATTTTTCATTTCATTTAATTCCTATATTCCTATTGATCCTATTACTTACTGTGCCT 2976

Qy   122 A 122
          |
Db   2975 A 2975

RESULT 5
AAZ80742
ID   AAZ80742 standard; cDNA; 507 BP.
XX
AC   AAZ80742;
XX
DT   07-APR-2000 (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:826.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia; hyperplasia;
KW   ds.
XX
OS   Homo sapiens.
XX

```
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;   99WO-IB001062.
XX
PR   10-JUN-1998;   98US-0088801P.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO,   Steinmann KE,   Astle JH,   Burgess CC,   Bushnell SE;
PI   Carroll E,   Catino TJ,   Derti A,   Ford DM,   Lewis ME,   Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer.
XX
PS   Claim 15; Page 459; Page 469pp; English.
XX
CC   AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which can
CC   be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC   and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC   for developing agents for the diagnosis and treatment of disorders
CC   involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC   hyperplasia
XX
SQ   Sequence 507 BP; 132 A; 148 C; 71 G; 152 T; 0 U; 4 Other;

Query Match             14.5%;   Score 35.6;  DB 3;  Length 507;
Best Local Similarity   62.2%;   Pred. No. 1.1;
Matches    56;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps     0;

Qy       62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA 121
            ||||||||||| ||||| || ||||  |||  |||| |||  ||  || |||  |||
Db      203 GATTTTAAATATACTACACCTCAGGGACCAAAAGAAAAAAGTAAGCAAGCAGGGGTTCCA 262

Qy      122 ACCTCTACCCCCCAACTGCAACGAGAGTT 151
            || |  |||||||  |||||||| |||||
Db      263 AGTGCTCTCCCCCAACTTCAACAAGAATGT 292
```

RESULT 6
AAL07156
```
ID   AAL07156 standard; DNA; 24387 BP.
XX
AC   AAL07156;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 9844.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184464P.
PR   02-MAR-2000; 2000US-0186350P.
```

```
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216447P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0254447P.
PR   14-AUG-2000; 2000US-0255757P.
PR   14-AUG-2000; 2000US-0255758P.
PR   14-AUG-2000; 2000US-0255759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   14-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236377P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
```

PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 9844; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 24387 BP; 5866 A; 6426 C; 6438 G; 5657 T; 0 U; 0 Other;

Query Match   14.5%; Score 35.6; DB 4; Length 24387;
Best Local Similarity   62.2%; Pred. No. 4.4;
Matches   56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTCTCCAGTCACCAGTGTCA 121
            |||||||||| ||||| ||| ||| ||| || |||||||||||||| || ||||||||
Db   19467 GATTTTAAATATACTACACCTCAGGGACCAAAAGAAAAAAGTTAAGCAAGCAGGGTTCCA 19526

Qy   122 ACCTCTACCCCCAACTGCAACGAGAGTTT 151
            || ||||||| ||| ||||||||| ||
Db   19527 AGTGCTCTCCCCAACTTCAACAAGAATGT 19556

RESULT 7
ABK35341/c
ID   ABK35341 standard; cDNA; 2803 BP.
XX
AC   ABK35341;
XX
DT   08-MAY-2002 (first entry)
XX
DE   Human cDNA encoding secreted protein #479.
XX
KW   Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW   viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW   autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW   autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW   Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW   coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW   tissue regeneration; wound healing; burn; haematopoiesis;
KW   myeloid cell deficiency; lymphoid cell deficiency.
XX
OS   Homo sapiens.
XX
PN   WO200177288-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US010224.
XX
PR   06-APR-2000; 2000US-0195582P.
XX
PA   (GEMY) GENETICS INST INC.
XX
PI   Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI   Gulukota K, Graham JR;
XX
DR   WPI; 2002-179321/23.
XX
PT   Five hundred and ninety two polynucleotides derived from a variety of
PT   human tissue sources which encode secreted proteins, useful for treating
PT   immune deficiencies and disorders such as autoimmune disorders.
XX
PS   Claim 1; Page 317-318; 372pp; English.
XX
CC   The invention relates to 592 polynucleotides which have been derived from
CC   a variety of human tissue sources and which encode novel secreted
CC   proteins. The polynucleotides can be used as probes for the
CC   identification and isolation of full length cDNA and genomic DNA. The
CC   polynucleotides and proteins can also be used as nutritional supplements.
CC   The proteins are useful in the treatment of various immune deficiencies
CC   and disorders such as viral infections, bacterial infections, fungal
CC   infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC   sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC   and conditions (e.g. asthma). They are also useful for treating
CC   neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
CC   disease), liver fibrosis, coagulation disorders (e.g. haemophilia,

CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC Sequences ABK34863 ABK35454 represent polynucleotides of the invention
XX
SQ Sequence 2803 BP; 688 A; 760 C; 712 G; 643 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 6; Length 2803;
Best Local Similarity 63.5%; Pred. No. 2.3;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy     2 TCATTAACCTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTATACACA   61
         |||||| |||||||||||||||| ||||||||| |||||||| | ||  |||||
Db  2102 TCATTTTCCCATCCCAGGATAGATATAATTTCTTTGATATTTATAAATATATATATA 2043

Qy    62 GATTTTAAGTTTACTCCTACTGCTG   86
         || |||  ||||
Db  2042 TATATTATATGTACACACACCTG   2018
```

RESULT 8
AAC84450/c
ID AAC84450 standard; cDNA; 3644 BP.
XX
AC AAC84450;
XX
DT 02-APR-2001 (first entry)
XX
DE Nucleotide sequence of mddt cDNA clone ID No: 244366.6.
XX
KW Molecule for disease detection and treatment; mddt; human; cirrhosis;
KW cell proliferative disorder; arteriosclerosis; bursitis; hepatitis;
KW connective tissue disease; myelofibrosis; cancer; leukemia; autoimmune;
KW acquired immunodeficiency syndrome; AIDS; Addison's disease; allergy;
KW anemia; bronchitis; gout; Hashimoto's thyroiditis; multiple sclerosis;
KW cytostatic; immunomodulator; anti-inflammatory; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200075298-A2.
XX
PD 14-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US015344.
XX
PR 03-JUN-1999; 99US-0137412P.
PR 05-AUG-1999; 99US-0147500P.
PR 05-AUG-1999; 99US-0147501P.
PR 05-AUG-1999; 99US-0147542P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM, Wright RJ;
PI Daniels SE;
XX
DR WPI; 2001-071068/08.
XX
PT New polypeptide and polynucleotide molecules for disease detection and
PT treatment are useful in diagnosis and gene therapy of proliferative
PT disorders (e.g. breast cancer), autoimmune and inflammatory disorders
PT (e.g. AIDS or allergy).
XX
PS Claim 1; Page 93-94; 99pp; English.
XX
CC Sequences AAC84445-C84458 are new isolated polynucleotide molecules for
CC disease detection and treatment (mddt). The mddt polynucleotides are
CC useful for diagnosing or treating disorders associated with disease
CC detection and treatment molecules These diseases include cell
CC proliferative disorders (e.g. arteriosclerosis, bursitis, cirrhosis,
CC hepatitis, mixed connective tissue disease, myelofibrosis, or cancers

CC such as leukemia, or breast or brain cancers), autoimmune or inflammatory
CC disorders (e.g. acquired immunodeficiency syndrome (AIDS), Addison's
CC disease, allergies, anemia, autoimmune haemolytic anemia, bronchitis,
CC gout, Hashimoto's thyroiditis, multiple sclerosis or cancers). The mddt
CC are especially useful for somatic or germline gene therapy. The mddt may
CC also be used to detect the presence of, or to quantify the amount of, an
CC mddt-related polynucleotide in a sample. The mddt are also useful for
CC isolating full length cDNA sequences utilizing hybridization and/or
CC amplification procedures, for generating hybridization probes useful in
CC chromosomal mapping of naturally occurring genomic sequences, as
CC molecular weight markers, or for monitoring the progress of disorders
CC associated with abnormal levels of mddt expression or evaluating the
CC efficacy of a particular treatment. The recombinant nucleic acids are
CC useful as part of a viral vector (e.g. based on a vaccinia virus) for
CC vaccinating a mammal and inducing a protective immunological response in
CC the mammal. The MDDT polypeptides are useful for screening molecules that
CC bind to or are bound by the polypeptides, i.e. for screening agonists,
CC antagonists or modulators of MDDT
XX
SQ Sequence 3644 BP; 856 A; 979 C; 897 G; 912 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 4; Length 3644;
Best Local Similarity 63.5%; Pred. No. 2.5;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy     2 TCATTAACCTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTATACACA   61
         |||||| |||||||||||||||| ||||||||| |||||||| | ||  |||||
Db  2938 TCATTTTCCCATCCCAGGATAGATATAATTTCTTTGATATTTATAAATATATATATA 2879

Qy    62 GATTTTAAGTTTACTCCTACTGCTG   86
         || |||  ||||
Db  2878 TATATTATATGTACACACACCTG   2854
```

RESULT 9
AAH17830/c
ID AAH17830 standard; cDNA; 3710 BP.
XX
AC AAH17830;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:17508.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 17508; 2537pp + Sequence Listing; English.

XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 3710 BP; 866 A; 999 C; 905 G; 940 T; 0 U; 0 Other;

Query Match   14.4%;   Score 35.4; DB 4;   Length 3710;
Best Local Similarity   63.5%;   Pred. No. 2.6;
Matches   54;   Conservative   0;   Mismatches   31;   Indels   0;   Gaps   0;

```
Qy      2 TCATTAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACA 61
            ||||||  ||||||||||||||||| || |||  || |||||| |||||||| |||||
Db   3009 TCATTTTCCATCCCAGGATAGATATATAAATTTCTTTGATATTTATAATATATATATATA 2950

Qy     62 GATTTTAAGTTTACTCCTACTGCTG 86
            |||||  |||| ||| ||||||||
Db   2949 TATATTATATGTACACACACACCTG 2925
```

RESULT 10
ABK28656/c
ID   ABK28656 standard; cDNA; 4480 BP.
XX
AC   ABK28656;
XX
DT   09-APR-2002   (first entry)
XX
DE   Human cDNA encoding secreted potein SECP23.
XX
KW   Human; ss; gene; SECP; antiinflammatory; cytostatic; cardiant;
KW   immunosuppressive; antiviral; anti-HIV; antiarthritic; antirheumatic;
KW   muscular active general; anticonvulsant; nootropic; neuroprotective;
KW   antiallergic; hypotensive; cardiovascular disorder; atherosclerosis;
KW   hypertension; myocardial infarction; autoimmune disorder;
KW   inflammatory disorder; AIDS; acquired immunodeficiency syndrome; allergy;
KW   rheumatoid arthritis; cell proliferative disorder; cancer;
KW   developmental disorder; Duchenne muscular dystrophy;
KW   neurological disorder; epilepsy; Alzheimer's disease.
XX
OS   Homo sapiens.
XX
PN   WO200198353-A2.
XX
PD   27-DEC-2001.
XX
PF   20-JUN-2001; 2001WO-US019862.
XX
PR   20-JUN-2000; 2000US-0212890P.
PR   23-JUN-2000; 2000US-0214466P.
PR   27-JUN-2000; 2000US-0214601P.
PR   31-JUL-2000; 2000US-0222372P.
PR   08-SEP-2000; 2000US-0231435P.
PR   15-SEP-2000; 2000US-0232889P.

XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,   Tang YT,   Yue H,   Elliott VS,   Tribouley CM,   Lee EA;
PI   Ramkumar JL,   Lal P,   Xu Y,   Warren BA,   Hafalia AJA,   Baughn MR;
PI   Azimzai Y,   Batra S,   Burford N,   Yao MG,   Nguyen DB,   Lu DAM,   Walia NK;
PI   Gandhi AR,   Au-Young J,   Patterson C;
XX
DR   WPI; 2002-090431/12.
DR   P-PSDB; AAU81997.
XX
PT   Forty four human secreted proteins (referred to as SECP-1 to SECP-44),
PT   useful in the diagnosis, treatment and prevention of cardiovascular (e.g.
PT   atherosclerosis), autoimmune/inflammatory (e.g. allergies) and cell
PT   proliferative disorders.
XX
PS   Claim 5; Page 180-182; 195pp; English.
XX
CC   The invention relates to forty four human secreted proteins (referred to
CC   as SECP-1 to SECP-44) and the nucleic acids encoding them. Also included
CC   are a host cell transformed with the nucleic acid, a transgenic animal
CC   comprising the nucleic acid, an anti-SECP antibody, use of the SECP
CC   proteins in isolating agonists and antagonists of SECP activity and a
CC   method of isolating compounds which alter the expression of the SECP
CC   nucleic acid. The SECP polynucleotides and polypeptides are useful in the
CC   diagnosis, treatment and prevention of cardiovascular (e.g.
CC   atherosclerosis, hypertension, myocardial infarction,
CC   autoimmune/inflammatory (e.g. acquired immunodeficiency syndrome (AIDS),
CC   allergies, rheumatoid arthritis), cell proliferative (e.g. cancer),
CC   developmental (e.g. Duchenne and Becker muscular dystrophy), and
CC   neurological (e.g. epilepsy, Alzheimer's disease) disorders. Numerous
CC   other examples of each disorder are given in the specification. The
CC   present sequence is a cDNA encoding a SECP protein
XX
SQ   Sequence 4480 BP; 1092 A; 1152 C; 1102 G; 1134 T; 0 U; 0 Other;

Query Match   14.4%;   Score 35.4.; DB 6;   Length 4480;
Best Local Similarity   63.5%;   Pred. No. 2.7;
Matches   54;   Conservative   0;   Mismatches   31;   Indels   0;   Gaps   0;

```
Qy      2 TCATTAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACA 61
            |||||| ||||||||||||||||| || |||  || |||||| |||||||| |||||
Db   3771 TCATTTTCCATCCCAGGATAGATATATAAATTTCTTTGATATTTATAATATATATATATA 3712

Qy     62 GATTTTAAGTTTACTCCTACTGCTG 86
            |||||  |||| ||| ||||||||
Db   3711 TATATTATATGTACACACACACCTG 3687
```

RESULT 11
ADE53917
ID   ADE53917 standard; cDNA; 2264 BP.
XX
AC   ADE53917;
XX
DT   29-JAN-2004   (first entry)
XX
DE   Human prostate cancer cDNA #264.
XX
KW   Human; prostate cancer; ss; cDNA combination; differential expression;
KW   gene.
XX
OS   Homo sapiens.
XX
PN   US2003190640-A1.
XX
PD   09-OCT-2003.
XX
PF   29-MAY-2002; 2002US-00252157.
XX
PR   31-MAY-2001; 2001US-0295048P.
XX
PA   (FARI/) FARIS M.

PA   (PEAR/) PEARSON C I.
XX
PI   Faris M,   Pearson CI;
XX
DR   WPI; 2003-831619/77.
XX
PT   New combination comprising cDNAs that are differentially expressed in
PT   prostate cancer, useful for diagnosing, treating or monitoring the
PT   progression of treatment of prostate cancer.
XX
PS   Claim 5; SEQ ID NO 264; 42pp; English.
XX
CC   The invention relates to a combination comprising a number of cDNAs
CC   expressed in prostate cancer. The invention also relates to a method for
CC   detecting differential expression of one or more cDNAs in a sample
CC   containing nucleic acids by hybridising a substrate with the nucleic
CC   acids, thus forming one or more hybridisation complexes, detecting
CC   hybridisation complex formation and comparing the complexes formed with.
CC   standard complexes, where differences between the standard and the sample
CC   complex formation indicate differential expression of cDNAs in the
CC   sample. The differential expression is diagnostic of prostate cancer. The
CC   invention also relates to proteins and antibodies related to the cDNAs.
CC   The combination is useful for diagnosing, treating or monitoring the
CC   progression of treatment of prostate cancer. The antibodies are useful
CC   for detecting prostate cancer. This sequence represents a human prostate
CC   cancer cDNA of the invention.
XX
SQ   Sequence 2264 BP; 472 A; 629 C; 585 G; 517 T; 0 U; 61 Other;

Query Match   14.1%; Score 34.8; DB 9; Length 2264;
Best Local Similarity 54.3%; Pred. No. 3.3;
Matches 69; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

Qy   25 TTTTTATTTCTTTAAGATTCCTGTGTTTATACACAGATTTAAGTTACTCCCTACTGC 84
            ||| |||| ||| ||||||||||||||||| |||| |||| |||||||||||||
Db   1452 TTCTGTTCGTTTTTTTTTTTTNCTTTTTTTTTCTATATATTTTTGTTGAATTCTATTTT 1511

Qy   85 TGACCCAAGTGAAATTCCTTCCCAGTCACAGTCAACCTCTACCCCCAACTGCAACG 144
            ||||||| ||||| ||| |||| || || ||| || ||||||| ||| |||
Db   1512 ATTTTTTAATTCTCTCTCTCTCTCTCCAGACACACAATGGCACTGCTTATCTCCGAAAATGGTGTG 1571

Qy   145 AGAGTTT 151
            || ||||
Db   1572 ATCGTCT 1578

RESULT 12
ABL03611/c
ID   ABL03611 standard; cDNA; 2239 BP.
XX
AC   ABL03611;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5315.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;

XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59508.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 5315; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16175-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2239 BP; 686 A; 496 C; 536 G; 521 T; 0 U; 0 Other;

Query Match   14.0%; Score 34.4; DB 4; Length 2239;
Best Local Similarity 63.1%; Pred. No. 4.3;
Matches 53; Conservative   0; Mismatches 31; Indels   0; Gaps   0;

Qy   32 TTTCTTTAAGATTCCTGTGTTTATACACAGATTTAAGTTACTCCTACTGCTGACCCA 91
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 TTTCCCTCACATCCATGTGTCTATATACAGTGTTCCATTTGACTTTGATCCTGACCTG 1519

Qy   92 AGTGAAATTCCTTCTCCAGTCACA 115
            ||||| |||||||| ||||||||
Db   1518 TTTGAAGATAATCCTCAACTCCCA 1495

RESULT 13
ABL03610/c
ID   ABL03610 standard; cDNA; 4405 BP.
XX
AC   ABL03610;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5312.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59507.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 5312; 21pp + Sequence Listing; English.
XX

CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 4405 BP; 1275 A; 956 C; 1086 G; 1088 T; 0 U; 0 Other;

Query Match   14.0%; Score 34.4; DB 4; Length 4405;
Best Local Similarity   63.1%; Pred. No. 5.5;
Matches   53;   Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   32 TTTCTTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTACTCCTACTGCTGACCCA 91
              |||| |||||||||||||||||||||||||||| ||||| ||||| |||||||||||||
Db   2744 TTTCCCTCACATCCATGTGTCTATATACAGTGTTCCATTTGACTTTTGATCCTGACCTG 2685

Qy   92 AGTGAAATTCCTTCTCCAGTCACA 115
              ||||| ||||||| ||||
Db   2684 TTTGAAGATAATCCTCAACTCCCA 2661

RESULT 14
ABL03530/c
ID   ABL03530 standard; cDNA; 12860 BP.
XX
AC   ABL03530;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 5072.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB59427.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 5072; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ   Sequence 12860 BP; 3591 A; 2899 C; 2995 G; 3375 T; 0 U; 0 Other;

Query Match   14.0%; Score 34.4; DB 4; Length 12860;
Best Local Similarity   63.1%; Pred. No. 8.1;
Matches   53;   Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   32 TTTCTTTAAGATTCCTGTTGTTTATACACAGATTTAAGTTACTCCTACTGCTGACCCA 91
              |||| |||||||||||||||||||||||||||| ||||| ||||| |||||||||||||
Db   10389 TTTCCCTCACATCCATGTGTCTATATACAGTGTTCCATTTGACTTTTGATCCTGACCTG 10330

Qy   92 AGTGAAATTCCTTCTCCAGTCACA 115
              ||||| ||||||| ||||
Db   10329 TTTGAAGATAATCCTCAACTCCCA 10306

RESULT 15
AAK55128
ID   AAK55128 standard; cDNA; 388 BP.
XX
AC   AAK55128;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:188.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ss.
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0252213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0277182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.

PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.

PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251866P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC,   Ruben SM;
XX
DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM82347.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Claim 1; SEQ ID NO 188; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 388 BP; 76 A; 131 C; 73 G; 105 T; 0 U; 3 Other;

Query Match   13.9%; Score 34.2; DB 4; Length 388;
Best Local Similarity   51.7%; Pred. No. 2.6;
Matches   78; Conservative   0; Mismatches   73; Indels   0; Gaps   0;

```
Qy    1 GTCATTAACCTTGCAAGGATACCTTTTATTTCTTAAGATTCCTGTTGTTTATACAC 60
         ||||  |||||| |||  ||||| |||    ||     ||| |||||| |||  ||||||
Db  205 GTCACAGTTCCTTCCCACCTGTGTTCTGTTTCGTTTTTTTTTCTTTTTTTTCTATAT 264

Qy   61 AGATTTTAAGTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTC 120
         |||||   || ||||| ||| |    |||||| ||||||||| ||||||||||||||||
Db  265 ATATTTTTGTTGAATTCTATTTTATTTTTAATTCCTCTCTCCTCCCCAGACACAATGGC 324

Qy  121 AACCTCTACCCCCAACTGCAACGAGAGTT 151
```

Db

325 ACTGCTTATCTCCGAAATGGTGTGTGATCGTCT 355

Search completed: June 10, 2004, 17:11:34
Job time : 364 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      June 10, 2004, 15:30:57 ; Search time 54 Seconds
                                       (without alignments)
                                       2528.110 Million cell updates/sec

Title:           US-09-674-593-10
Perfect score:   246
Sequence:        1 gtcattaacctttgcaagga.........ttcatatcaacccacacta 246

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        682709 seqs, 277747446 residues

Total number of hits satisfying chosen parameters:        1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the total of the result being printed,
         and is derived by analysis of the total score distribution.

                                SUMMARIES
```

Result No.	Score	Query Match %	Length	DB	ID	Description
1	35.6	14.5	507	3	US-09-328-111-826	Sequence 826, App
c 2	34	13.8	786431	4	US-09-751-389-3	Sequence 3, Appli
c 3	31.8	12.9	1885	4	US-09-023-655-1162	Sequence 1162, Ap
4	31.6	12.8	2337	4	US-09-328-352-3781	Sequence 3781, Ap
5	31.6	12.8	13158	2	US-08-687-080-105	Sequence 105, App
6	31.4	12.8	1017	4	US-09-601-198-42	Sequence 42, Appl
c 7	31.2	12.7	530	4	US-09-389-681-434	Sequence 434, App
c 8	31.2	12.7	530	4	US-09-620-405B-434	Sequence 434, App
c 9	31.2	12.7	530	4	US-09-433-826B-434	Sequence 434, App
c 10	31.2	12.7	530	4	US-09-604-287A-434	Sequence 434, App
c 11	31.2	12.7	530	4	US-09-834-759-434	Sequence 434, App
c 12	31.2	12.7	704	3	US-08-235-836C-125	Sequence 125, App
13	31	12.6	704	3	US-08-235-836C-126	Sequence 126, App
c 14	31	12.6	1413	4	US-09-342-459-1	Sequence 1, Appli
c 15	31	12.6	3931	4	US-08-956-171E-342	Sequence 342, App
16	31	12.6	5184	4	US-09-976-594-484	Sequence 484, App
c 17	30.6	12.4	550	4	US-09-621-976-3345	Sequence 3345, Ap
18	30.6	12.4	741	4	US-09-621-976-1894	Sequence 1894, Ap
c 19	30.2	12.3	9278	1	US-08-243-542-9	Sequence 9, Appli
20	30.2	12.3	9278	1	US-08-477-407-9	Sequence 9, Appli
21	30.2	12.3	9278	1	US-08-484-355-9	Sequence 9, Appli
c 22	30	12.2	129908	4	US-09-585-858-1	Sequence 1, Appli
23	29.8	12.1	8920	2	US-08-446-855A-1	Sequence 1, Appli
24	29.8	12.1	8920	3	US-09-150-741-1	Sequence 1, Appli
25	29.6	12.0	1704	4	US-09-023-655-263	Sequence 263, App
c 26	29.6	12.0	3018	4	US-09-206-942-40	Sequence 40, Appl
c 27	29.6	12.0	3036	4	US-09-206-942-38	Sequence 38, Appl
c 28	29.6	12.0	3966	3	US-09-215-131-1	Sequence 1, Appli
c 29	29.6	12.0	3966	3	US-09-222-734-1	Sequence 1, Appli
c 30	29.4	12.0	324	3	US-09-107-532A-3015	Sequence 3015, Ap
c 31	29.4	12.0	704	3	US-08-235-836C-123	Sequence 123, App
c 32	29.4	12.0	704	3	US-08-235-836C-124	Sequence 124, App
33	29.2	11.9	337	4	US-08-956-171E-1348	Sequence 1348, Ap
c 34	28.8	11.7	495	4	US-09-543-681A-2486	Sequence 2486, Ap
35	28.6	11.6	503	3	US-09-621-976-15042	Sequence 15042, A
36	28.6	11.6	858	4	US-09-328-352-3863	Sequence 3863, Ap
37	28.6	11.6	1425	2	US-09-134-001C-404	Sequence 404, App
c 38	28.6	11.6	1882	3	US-08-867-970-1	Sequence 1, Appli
c 39	28.6	11.6	1882	3	US-09-326-217-1	Sequence 1, Appli
c 40	28.6	11.6	1882	4	US-09-732-020-1	Sequence 1, Appli
41	28.6	11.6	2946	3	US-08-968-563-6	Sequence 6, Appli
42	28.6	11.6	2946	3	US-08-969-683A-6	Sequence 6, Appli
43	28.6	11.6	2946	4	US-09-297-928-2	Sequence 2, Appli
44	28.4	11.5	56516	2	US-08-996-306-1	Sequence 1, Appli
45	28.4	11.5	56516	3	US-09-338-907-1	Sequence 1, Appli

```
                                ALIGNMENTS

RESULT 1
US-09-328-111-826
; Sequence 826, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 826
;  LENGTH: 507
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(507)
;  OTHER INFORMATION: n = A,T,C or G
US-09-328-111-826

Query Match          14.5%;   Score 35.6;   DB 3;   Length 507;
Best Local Similarity 62.2%;   Pred. No. 0.044;
Matches  56;  Conservative  0;  Mismatches  34;  Indels  0;  Gaps  0;

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCA 121
         |||||||||||||  |||||  ||  |||| ||  |||||||||||  ||||| ||||||
Db  203 GATTTTAAATATCATACACCTTGAGGACCCAGGACCAAAGAGAAAAAGTAAGCAAGCAAGCAGGGTTCCA 262

Qy  122 ACCTCTCACCCCCCAACTGCAACGACGAGTT 151
          |||||| |||||||||  ||| ||||| ||
Db  263 AGTGCTCCTCCCCCAACTTCAACAAGAATGT 292

RESULT 2
```

US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match   13.8%; Score 34; DB 4; Length 786431;
Best Local Similarity   74.1%; Pred. No. 4.5;
Matches   43; Conservative   0; Mismatches   15; Indels   0; Gaps   0;

Qy   16 AAGGATACCTTTTATTTTCTTTAAGATTCCTGTGTTTATACACAGATTTTAAGTTT 73
              ||||||| ||||||| ||||| | || ||||| ||||||||||| ||
Db   691859 ATGGATACATTTTTATATATATTTAAAATTTTTATCTTTAAACACCACTTTAAGATT 691802

RESULT 3
US-09-023-655-1162
; Sequence 1162, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g187386
US-09-023-655-1162

Query Match   12.9%; Score 31.8; DB 4; Length 1885;
Best Local Similarity   53.7%; Pred. No. 1.4;
Matches   66; Conservative   0; Mismatches   57; Indels   0; Gaps   0;

Qy   29 TATTTCTTAAGATTCCTGTGTTTATACACAGATTTTAAGTTTACTCCTACTGCTGAC 88
              |||||| |||| | || ||||||||| ||||
Db   1426 TGTCTTTTTTTTATTTTACTTTTTTTTAAGCACCAAAATTTGTGTTTTTTTTCTCCCC 1485

Qy   89 CCAAGTGAAATTCCTTCCCAGTCACAGTGTCAACCTCTACCCCCCAACTGCAACGAGAG 148
                           |||| ||| ||| |||| |||||||||||||||||
Db   1486 TCCCCACAGATCCCATCCAAATCATTCGTTAACCACCATTCCAACAGGTCGAGGAGAG 1545

Qy   149 TTT 151
              |||
Db   1546 CTT 1548

RESULT 4
US-09-328-352-3781
; Sequence 3781, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3781
; LENGTH: 2337
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3781

Query Match   12.8%; Score 31.6; DB 4; Length 2337;
Best Local Similarity   51.4%; Pred. No. 1.8;
Matches   73; Conservative   0; Mismatches   69; Indels   0; Gaps   0;

Qy   13 TGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTGTTTATACACAGATTTTAAGTT 72
              ||||| ||||||| || ||||| ||||| |||||||||||||
Db   1281 TGAGACGGGAAGGTTTTGATCAGCTTCAAGACTTTGAAGGTTTTGAAAAGATTTATGTCTT 1340

Qy   73 TACTCCTACTGCTGACCCAAGTGAAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCC 132
                  ||||||| |||| ||| | |||| ||||| |||||||||||
Db   1341 TAGCCGTGATGTGATATTAAGAACTACCACGTGGCTCAACAGTTTTAGACTTTGCTTA 1400

Qy   133 CCAACTGCAACGAGAGTTTGA 154
              |||| ||||||||| ||
Db   1401 CCATGTGCATACAGAGGTTGGA 1422

RESULT 5
US-08-687-080-105
; Sequence 105, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRES:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA

```
;  COUNTRY: USA
;  ZIP: 94306
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/687,080
;  FILING DATE: 17-JUL-1996
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US 08/592,126
;  FILING DATE: 26-JAN-1996
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Sholtz, Charles K.
;  REGISTRATION NUMBER: 38,615
;  REFERENCE/DOCKET NUMBER: 4600-0111.30
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (415) 324-0880
;  TELEFAX: (415) 324-0960
;  INFORMATION FOR SEQ ID NO: 105:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 13158 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: double
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
;  HYPOTHETICAL: NO
;  ANTI-SENSE: NO
;  ORIGINAL SOURCE:
;  INDIVIDUAL ISOLATE: 5' END OF INTRON 21 OF RAD50 GENOMIC
;  INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-105
```

```
Query Match       12.8%;   Score 31.6;   DB 2;   Length 13158;
Best Local Similarity   69.4%;   Pred. No. 4;
Matches  43;  Conservative  0;  Mismatches  19;  Indels  0;  Gaps  0;

Qy      25 TTTTTATTTTCTTTAAGATTCCTGTGTTGTTATACACAGATTTTAAGTTTACTCCCTACTGC 84
           |||||| ||||||||||||||     |||| ||    |||||| |||||| ||||||||||||
Db    4363 TTTATATTCTCTTTCAGATTGTTATTTTTTAATTCACATTTGAGGAGTCTGCTTCTGC 4422

Qy      85 TG 86
           ||
Db    4423 TG 4424
```

```
RESULT 6
US-09-601-198-42
;  Sequence 42, Application US/0960198
;  Patent No. 6531583
;  GENERAL INFORMATION:
;  APPLICANT: Cassell, Gail H.
;  APPLICANT: Chen, Ellson Y.
;  APPLICANT: Glass, Jennifer S.
;  APPLICANT: Glass, John I.
;  APPLICANT: Heiner, Cheryl R.
;  APPLICANT: Lefkowitz, Elliot
;  TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
;  TITLE OF INVENTION: UREALYTICUM
;  FILE REFERENCE: UAB-13452/22
;  CURRENT APPLICATION NUMBER: US/09/601,198
;  CURRENT FILING DATE: 2000-12-08
;  PRIOR APPLICATION NUMBER: 60/073,189
;  PRIOR FILING DATE: 1998-01-30
;  NUMBER OF SEQ ID NOS: 181
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 42
;  LENGTH: 1017
;  TYPE: DNA
;  ORGANISM: Ureaplasma urealyticum
US-09-601-198-42
```

```
Query Match       12.8%;   Score 31.4;   DB 4;   Length 1017;
Best Local Similarity   53.7%;   Pred. No. 1.4;
Matches  65;  Conservative  0;  Mismatches  56;  Indels  0;  Gaps  0;

Qy       7 AACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTGTTGTTATACACAGATTT 66
           ||||||| || |||||  ||||||| || || |||||||| |||||| | || ||| |||
Db     589 AATATTCTCCTGTATAAATTCTTCATTGCTTTGATTTTCGTTCTTTATAGGCTCAGGTTC 648

Qy      67 TAAGTTACTCCTACTGCTGACCCAGTGAAATTCCTTCTCCAGTGCACAGTGTCAACCTC 126
           || |||||| | |||  || |||| | |||  |||||| | |||| || |||||||||||
Db     649 AACTTCTTCATGTAATGGTTCAATAGGTGTAGATTCATGTTCGTTAACAATATCATCTTC 708

Qy     127 T 127
           |
Db     709 T 709
```

```
RESULT 7
US-09-389-681-434/c
;  Sequence 434, Application US/09389681A
;  Patent No. 6518237
;  GENERAL INFORMATION:
;  APPLICANT: Yuqui, Jiang
;  APPLICANT: Dillon, Davin C.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: Xu, Jiangchun
;  TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
;  TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.470C3
;  CURRENT APPLICATION NUMBER: US/09/389,681A
;  CURRENT FILING DATE: 1999-09-02
;  NUMBER OF SEQ ID NOS: 463
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 434
;  LENGTH: 530
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-389-681-434
```

```
Query Match       12.7%;   Score 31.2;   DB 4;   Length 530;
Best Local Similarity   60.7%;   Pred. No. 1.2;
Matches  51;  Conservative  0;  Mismatches  33;  Indels  0;  Gaps  0;

Qy      24 CTTTTATTTTCTTTAAGATTCCTGTGTTGTTATACACAGATTTAAGTTTACTCCTACTG 83
           ||||||||| ||||||||||||  ||| |||| || || || |||||| ||||||||||
Db     466 CTTTTTAATTCTTTAACTTTACTTTTTCTCATAGAGAGAGTCTTCGTTCATTATCCTCACT 407

Qy      84 CTGACCCAAGTGAAATTCCTTCTC 107
           ||||||| || ||| |||||||||
Db     406 CTCACAAAATCAGAATCATCCTC 383
```

```
RESULT 8
US-09-620-405B-434/c
;  Sequence 434, Application US/0962040405B
;  Patent No. 6528054
;  GENERAL INFORMATION:
;  APPLICANT: Jiang, Yuqiu
;  APPLICANT: Dillon, Davin C.
;  APPLICANT: Mitcham, Jennifer L.
;  APPLICANT: Xu, Jiangchun
;  APPLICANT: Harlocker, Susan L.
;  APPLICANT: Hepler, William T.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
;  FILE REFERENCE: 210121.470C8
;  CURRENT APPLICATION NUMBER: US/09/620,405B
;  CURRENT FILING DATE: 2000-07-20
;  NUMBER OF SEQ ID NOS: 495
;  SOFTWARE: FastSEQ for Windows Version 3.0
;  SEQ ID NO 434
;  LENGTH: 530
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-620-405B-434

Query Match      12.7%;   Score 31.2;   DB 4;   Length 530;
Best Local Similarity  60.7%;   Pred. No. 1.2;
Matches  51;   Conservative  0;   Mismatches  33;   Indels  0;   Gaps  0;

Qy    24  CTTTTTATTTCTTTAAGATTCCTGTGTTGTTTATACACAGATTTAAGTTTACTCCTACTG   83
          ||||||  ||||||||     |||  |||  ||||||||||||||||| || |  |||||
Db   466  CTTTTTAATTTCTTTAACTTTACTTTTTCTCATAGAGAAGTCTTCGTCATTATCCTCACT  407

Qy    84  CTGACCCAAGTGAAATTCCTTCTC  107
          ||| ||||||||||| |||||
Db   406  CTCACAAAAATCAGAATCATCCTC  383
```

RESULT 9
US-09-433-826B-434/c
```
;   Sequence 434, Application US/09433826B
;   Patent No. 6579973
;   GENERAL INFORMATION:
;   APPLICANT: Jiang, Yuqui
;   APPLICANT: Dillon, Davin C.
;   APPLICANT: Mitcham, Jennifer L.
;   APPLICANT: Xu, Jiangchun
;   APPLICANT: Harlocker, Susan L.
;   APPLICANT: Hepler, William T.
;   TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
;   TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
;   FILE REFERENCE: 210121.470C4
;   CURRENT APPLICATION NUMBER: US/09/433,826B
;   CURRENT FILING DATE: 1999-11-03
;   NUMBER OF SEQ ID NOS: 474
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 434
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-433-826B-434

Query Match      12.7%;   Score 31.2;   DB 4;   Length 530;
Best Local Similarity  60.7%;   Pred. No. 1.2;
Matches  51;   Conservative  0;   Mismatches  33;   Indels  0;   Gaps  0;

Qy    24  CTTTTTATTTCTTTAAGATTCCTGTGTTGTTTATACACAGATTTAAGTTTACTCCTACTG   83
          ||||||  ||||||||     |||  |||  ||||||||||||||||| || |  |||||
Db   466  CTTTTTAATTTCTTTAACTTTACTTTTTCTCATAGAGAAGTCTTCGTCATTATCCTCACT  407

Qy    84  CTGACCCAAGTGAAATTCCTTCTC  107
          ||| ||||||||||| |||||
Db   406  CTCACAAAAATCAGAATCATCCTC  383
```

RESULT 10
US-09-604-287A-434/c
```
;   Sequence 434, Application US/09604287A
;   Patent No. 6586572
;   GENERAL INFORMATION:
;   APPLICANT: Jiang, Yuqui
;   APPLICANT: Dillon, Davin C.
;   APPLICANT: Mitcham, Jennifer L.
;   APPLICANT: Xu, Jiangchun
;   APPLICANT: Harlocker, Susan L.
;   APPLICANT: Hepler, William T.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;   TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
;   FILE REFERENCE: 210121.470C7
;   CURRENT APPLICATION NUMBER: US/09/604,287A
;   CURRENT FILING DATE: 2000-06-22
;   NUMBER OF SEQ ID NOS: 489
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 434
;   LENGTH: 530
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-604-287A-434

Query Match      12.7%;   Score 31.2;   DB 4;   Length 530;
Best Local Similarity  60.7%;   Pred. No. 1.2;
Matches  51;   Conservative  0;   Mismatches  33;   Indels  0;   Gaps  0;

Qy    24  CTTTTTATTTCTTTAAGATTCCTGTGTTGTTTATACACAGATTTAAGTTTACTCCTACTG   83
          ||||||  ||||||||     |||  |||  ||||||||||||||||| || |  |||||
Db   466  CTTTTTAATTTCTTTAACTTTACTTTTTCTCATAGAGAAGTCTTCGTCATTATCCTCACT  407

Qy    84  CTGACCCAAGTGAAATTCCTTCTC  107
          ||| ||||||||||| |||||
Db   406  CTCACAAAAATCAGAATCATCCTC  383
```

RESULT 11
US-09-834-759-434/c
```
;   Sequence 434, Application US/09834759
;   Patent No. 6680197
;   GENERAL INFORMATION:
;   APPLICANT: Jiang, Yuqiu
;   APPLICANT: Dillon, Davin C.
;   APPLICANT: Mitcham, Jennifer L.
;   APPLICANT: Xu, Jiangchun
;   APPLICANT: Harlocker, Susan L.
;   APPLICANT: Hepler, William T.
;   APPLICANT: Henderson, Robert A.
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;   TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
;   FILE REFERENCE: 210121.470C9
;   CURRENT APPLICATION NUMBER: US/09/834,759
;   CURRENT FILING DATE: 2001-04-13
;   NUMBER OF SEQ ID NOS: 547
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 434
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-834-759-434

Query Match      12.7%;   Score 31.2;   DB 4;   Length 530;
Best Local Similarity  60.7%;   Pred. No. 1.2;
Matches  51;   Conservative  0;   Mismatches  33;   Indels  0;   Gaps  0;

Qy    24  CTTTTTATTTCTTTAAGATTCCTGTGTTGTTTATACACAGATTTAAGTTTACTCCTACTG   83
          ||||||  ||||||||     |||  |||  ||||||||||||||||| || |  |||||
Db   466  CTTTTTAATTTCTTTAACTTTACTTTTTCTCATAGAGAAGTCTTCGTCATTATCCTCACT  407

Qy    84  CTGACCCAAGTGAAATTCCTTCTC  107
          ||| ||||||||||| |||||
Db   406  CTCACAAAAATCAGAATCATCCTC  383
```

RESULT 12
US-08-235-836C-125/c
```
;   Sequence 125, Application US/08235836C
;   Patent No. 6248562
;   GENERAL INFORMATION:
;   APPLICANT: Dunn, John J.
;   APPLICANT: Luft, Benjamin J.
;   TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
;   TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
;   NUMBER OF SEQUENCES: 144
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brookhaven National Laboratory
;   STREET: Upton
;   CITY: Upton
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 11973
;   COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/235,836C
  FILING DATE: 29-APR-1994
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/148,191
  FILING DATE: 01-11-93
ATTORNEY/AGENT INFORMATION:
  NAME: Bogosian, Margaret C.
  REGISTRATION NUMBER: 25,324
  REFERENCE/DOCKET NUMBER: BNL93-28A
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 282-7338
  TELEFAX: (516) 282-3729
INFORMATION FOR SEQ ID NO: 125:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 704 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
US-08-235-836C-125

Query Match   12.6%;   Score 31; DB 3; Length 704;
Best Local Similarity   52.8%;   Pred. No. 1.6;
Matches 67; Conservative   0; Mismatches 60; Indels   0; Gaps   0;

Qy   2 TCATTAACCTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACACA 61
           || ||| |||| ||| ||| |||| ||| ||| |||| ||||| |||||| ||||
Db   463 TCGTTCTGCTTCATCAGAATCAGATAGTGTTTTAGTTGCTTCTTCTTGCTCTTTGTACAGT 404

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCCCAGTCACAGTGTCA 121
           ||||| |||||| |||||||
Db   403 TTTTGCATATTTGTTCTCACTGCTTTAGTAAGGTTATATGCTTCTTTCGCAACATTAGCC 344

Qy   122 ACCTCTA 128
           ||| |||
Db   343 GCAACTA 337

US-08-235-836C-126

INFORMATION FOR SEQ ID NO: 126:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 704 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 282-7338
  TELEFAX: (516) 282-3729
  REFERENCE/DOCKET NUMBER: BNL93-28A
  REGISTRATION NUMBER: 25,324
NAME: Bogosian, Margaret C.

Query Match   12.6%;   Score 31; DB 3; Length 704;
Best Local Similarity   52.8%;   Pred. No. 1.6;
Matches 67; Conservative   0; Mismatches 60; Indels   0; Gaps   0;

Qy   2 TCATTAACCTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACACA 61
           || ||| |||| ||| ||| |||| ||| ||| |||| ||||| |||||| ||||
Db   463 TCGTTCTGCTTCATCAGAATCAGATAGTGTTTTAGTTGCTTCTTCTTGCTCTTTGTACAGT 404

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTTCCCAGTCACAGTGTCA 121
           ||||| |||||| |||||||
Db   403 TTTTGCATATTTGTTCTCACTGCTTTAGTAAGGTTATATGCTTCTTTCGCAACATTAGCC 344

Qy   122 ACCTCTA 128
           ||| |||
Db   343 GCAACTA 337

RESULT 14
US-09-342-459-1/c
; Sequence 1, Application US/09342459
; Patent No. 6251633
; GENERAL INFORMATION:
;   APPLICANT: Yan, Kang
;   APPLICANT: Pearce, Kenneth
;   TITLE OF INVENTION: FtsA
;   FILE REFERENCE: GM10229
;   CURRENT APPLICATION NUMBER: US/09/342,459
;   CURRENT FILING DATE: 1999-06-29
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1413
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-342-459-1

Query Match   12.6%;   Score 31; DB 3; Length 1413;
Best Local Similarity   56.3%;   Pred. No. 2.2;
Matches 58; Conservative   0; Mismatches 45; Indels   0; Gaps   0;

Qy   2 TCATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTTATACACA 61
           | ||| |||| ||| || ||| |||| ||| ||| |||| ||||| |||||| ||||
Db   1364 TCTTTATCTTTGCTTGTGAACATGTTCATGATTCGTGTTTATGTTCCTGATGATTATCTTCT 1305

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTT 104
           ||||| |||||| |||||||
Db   1304 GATTGATAAATTTCTTCATCGTTGACTCTACTTCATTTTCAT 1262

RESULT 15
US-08-956-171E-342/c
; Sequence 342, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   Gil H. Choi
;   Patrick S. Dillon
;   Craig A. Rosen
;   Steven C. Barash

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/235,836C
  FILING DATE: 29-APR-1994
  CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 08/148,191
  FILING DATE: 01-11-93
ATTORNEY/AGENT INFORMATION:

RESULT 13
US-08-235-836C-126/c
; Sequence 126, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
;   APPLICANT: Dunn, John J.
;   APPLICANT: Luft, Benjamin J.
;   TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
;   TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
;   NUMBER OF SEQUENCES: 144
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brookhaven National Laboratory
;   STREET: Upton
;   CITY: Upton
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 11973
;   COMPUTER READABLE FORM:

Michael R. Fannon
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5256
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Human Genome Sciences, Inc.
   STREET: 9410 Key West Avenue
   CITY: Rockville
   STATE: Maryland
   COUNTRY: USA
   ZIP: 20850
COMPUTER READABLE FORM:
   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
   COMPUTER: HP Vectra 486/33
   OPERATING SYSTEM: MSDOS version 6.2
   SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/08/956,171E
   FILING DATE: 20-Oct-1997
   CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: 60/009,861
   FILING DATE: January 5, 1996
   APPLICATION NUMBER: 08/781,986
   FILING DATE: January 3, 1997
ATTORNEY/AGENT INFORMATION:
   NAME: Mark J. Hyman
   REGISTRATION NUMBER: 46,789
   REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (240) 314-1224
   TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 342:
SEQUENCE CHARACTERISTICS:
   LENGTH: 3931 base pairs
   TYPE: nucleic acid
   STRANDEDNESS: double
   TOPOLOGY: linear
   SEQUENCE DESCRIPTION: SEQ ID NO: 342:
US-08-956-171E-342

Query Match   12.6%; Score 31; DB 4; Length 3931;
Best Local Similarity   56.3%; Pred. No. 3.6;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy   2 TCATTAACCTTTGCAAGGATACCTTTTTATTTCTTTAAGATTCCTGTTGTTTATACACA 61
            || |||| ||| || ||| | || || ||| || | || |||| ||||| | |
Db   500 TCTTTATCTTTGTCTGAACATGTTCATGATTCTGTTTATGTTCCTGATGATTATCTTCT 441

Qy   62 GATTTTAAGTTTACTCCTACTGCTGACCCAAGTGAAATTCCTT 104
            ||| | |||| | ||| | || | || | || |||| |
Db   440 GATTGATAAATTCTTCATCTGTTGACTCTCACTTCATTTTCAT 398

Search completed: June 10, 2004, 17:01:40
Job time : 59 secs